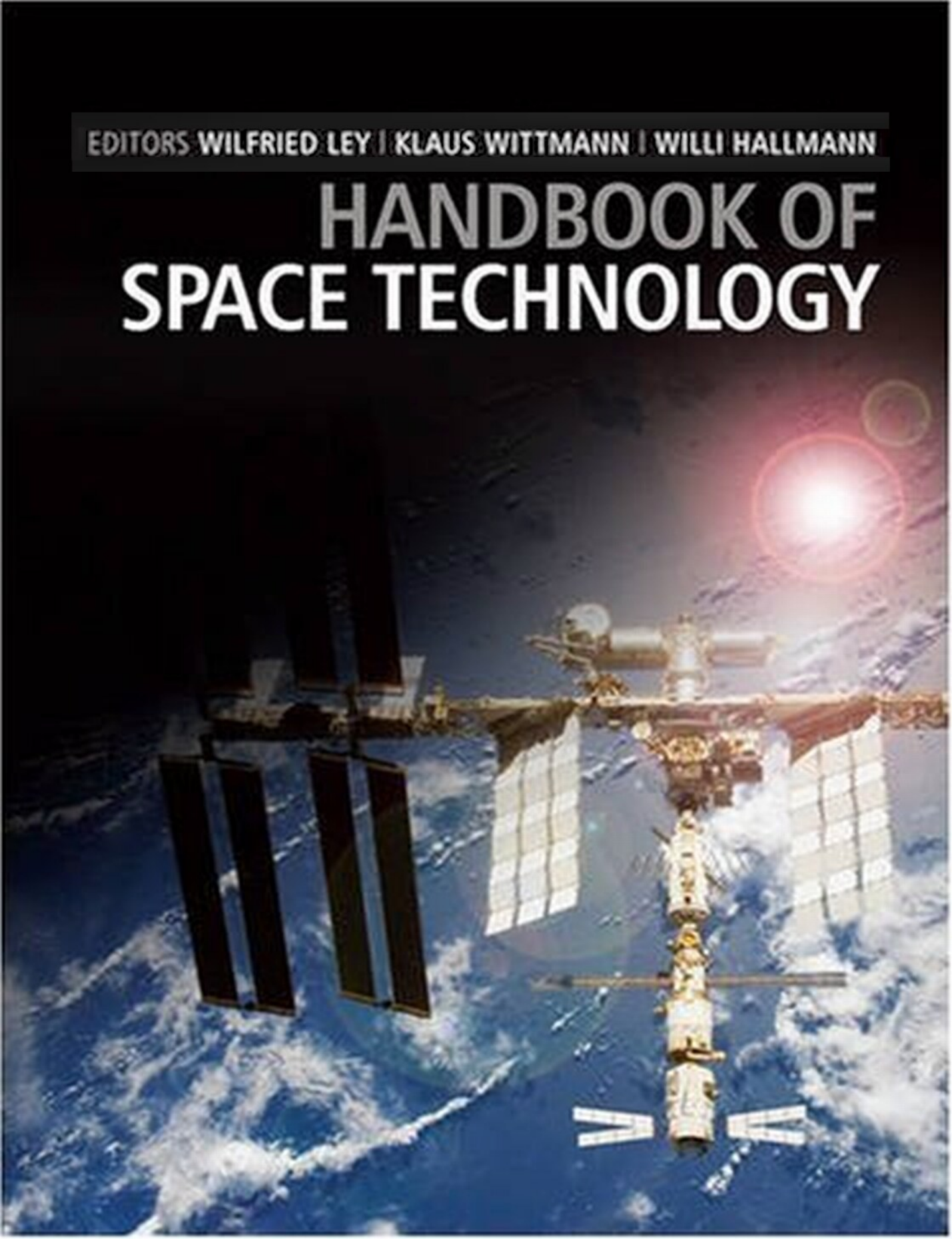
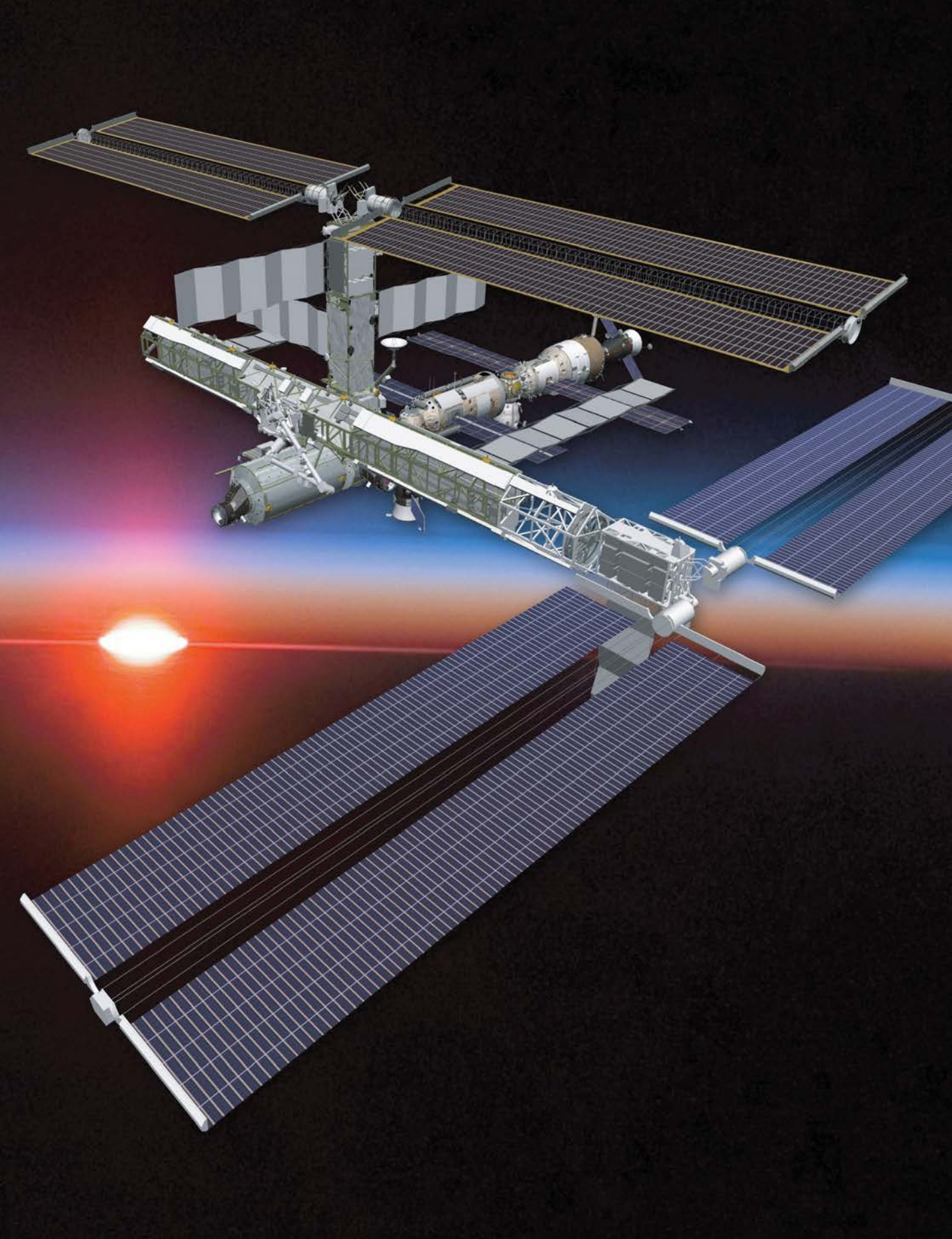
EDITORS WILFRIED LEY | KLAUS WITTMANN | WILLI HALLMANN

HANDBOOK OF SPACE TECHNOLOGY



Handbook of Space Technology

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9



Wilfried Ley/Klaus Wittmann/Willi Hallmann (Editors)

Handbook of Space Technology



A John Wiley and Sons, Ltd., Publication

This edition first published 2009 by John Wiley & Sons, Ltd

First published under the title Handbuch der Raumfahrttechnik 3.A.
by Carl Hanser Verlag © 2008
Editors: Wilfried Ley, Klaus Wittmann, Willi Hallmann
Carl Hanser Verlag GmbH & Co. KG, Munich/FRG, 2008.
All rights reserved

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the authors to be identified as the authors of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Record on File

ISBN: 978-0-470-69739-9

A catalogue record for this book is available from the British Library.

Typeset in 10/12pt Minion by Laserwords Private Limited, Chennai, India
Printed in Singapore by Markono

Contents

Foreword xv

Preface xvii

The Editors xix

The Authors xix

1 Introduction 1
Bibliography 4

1.1 Historical Overview 4
1.1.1 Introduction 4
1.1.2 The Development of Unmanned German and European Space Flight 6
1.1.3 The Development of Human Space Flight in Europe 9
Bibliography 15

1.2 Space Missions 16
1.2.1 Space System Segments 16
1.2.2 Design of System Segments for Space Flight Missions 22
1.2.3 Space Flight Mission Classification 25
Bibliography 29

2 Fundamentals 33

2.1 The Space Environment 33
2.1.1 Spacecraft and the Space Environment 33
2.1.2 Influence of the Sun and the Space Background 35
2.1.3 Influence of the Earth 39
2.1.4 Effect on Spacecraft and Mission Design 43
Bibliography 51

2.2 Orbital Mechanics 52
2.2.1 Orbit Modeling 52
2.2.2 Orbit Determination 66
2.2.3 Orbit Design and Station Keeping 74
Bibliography 81

2.3 Aerothermodynamics and Reentry 82
2.3.1 Introduction 82
2.3.2 Global Energy Considerations 83
2.3.3 Fluid Mechanical and Chemical Phenomena during Reentry 84
2.3.4 Heat Flux Balance and Thermal Protection Systems 87

2.3.5 Reentry Trajectory . . . 90
2.3.6 Aerodynamic Considerations . . . 92
2.3.7 Tools for the Determination of Aerothermodynamic Data . . . 95
Bibliography . . . 99

2.4 Meteoroids and Space Debris . . . 100
2.4.1 The Environmental Conditions . . . 100
2.4.2 Future Development and Debris Mitigation Measures . . . 101
2.4.3 Impact Flux and Impact Risk . . . 103
2.4.4 Protection of Spacecraft Against Impacting Particles . . . 107
2.4.5 Mission Planning . . . 112
Bibliography . . . 112

3 Space Transportation Systems . . . 115

3.1 Systems . . . 116
3.1.1 Introduction . . . 116
3.1.2 Fundamentals . . . 118
3.1.3 Building Blocks . . . 121
3.1.4 Project Phases . . . 121
3.1.5 Overview of Launch Systems . . . 122
Bibliography . . . 135

3.2 Multistage Rocket Technologies . . . 135
3.2.1 Introduction and Overview . . . 136
3.2.2 Mission Profiles and Operation . . . 139
3.2.3 Components and Subsystems . . . 142
3.2.4 Stage System Design Process and Technology . . . 149

3.3 Propulsion Systems . . . 154
3.3.1 Chemical Propulsion Basics . . . 155
3.3.2 Types of Engines . . . 156
3.3.3 Engine Components . . . 163
3.3.4 Special Problems . . . 178
3.3.5 Facilities for Rocket Engine Testing . . . 180
3.3.6 Future Propulsion Systems . . . 181
Bibliography . . . 183

3.4 Launch Infrastructure . . . 184
3.4.1 Requirements and Missions . . . 184
3.4.2 Concepts . . . 184
3.4.3 One Realized Example: Ariane 5 . . . 185
3.4.4 Major Launch Sites . . . 189

3.5 System Qualification . . . 189
3.5.1 Introduction . . . 189
3.5.2 Categories of Qualification . . . 195

3.5.3 Mechanical Qualification . . . 195
3.5.4 Functional Qualification . . . 197

4 Subsystems of Spacecraft . . . 201

4.1 Structure and Mechanisms . . . 203
4.1.1 The Primary Structure of the Spacecraft . . . 203
4.1.2 Secondary and Deployable Structures . . . 214
4.1.3 Structural Analysis . . . 217
4.1.4 Qualification of the Spacecraft Structure . . . 222
4.1.5 Mechanisms . . . 223
Bibliography . . . 235

4.2 Electrical Power Supply . . . 236
4.2.1 Energy Generation . . . 237
4.2.2 Power Sources . . . 237
4.2.3 Designing an Optimized Electrical Power System Architecture . . . 243
4.2.4 Electrical Power System Architectures . . . 245
4.2.5 Solar Array . . . 249
4.2.6 Energy Storage . . . 257
4.2.7 Design Fundamentals of EPS Systems . . . 264
Bibliography . . . 267

4.3 Thermal Control . . . 268
4.3.1 Introduction . . . 268
4.3.2 Basic Thermal Principles . . . 268
4.3.3 Development of the Thermal System . . . 276
4.3.4 Technical Solutions . . . 283
4.3.5 Example of a Thermal Design . . . 292
4.3.6 Operation of the Thermal Control System . . . 296
Bibliography . . . 301

4.4 Satellite Propulsion . . . 301
4.4.1 Fundamentals of Satellite Propulsion . . . 301
4.4.2 Propulsion System Types . . . 302
4.4.3 Propellants . . . 305
4.4.4 Feed Systems and Propellant Storage . . . 307
4.4.5 Cold Gas Propulsion Systems . . . 313
4.4.6 Chemical Propulsion . . . 315
4.4.7 Electric Propulsion . . . 324
4.4.8 Components for Chemical Propulsion Systems . . . 328
4.4.9 Ground Support Equipment and Services . . . 330
Bibliography . . . 331

4.5 Attitude Control . . . 332
4.5.1 Introduction and Overview . . . 332
4.5.2 Requirements for the Attitude Control System . . . 333

4.5.3 Parameterization of the Attitude . . . 334
4.5.4 Attitude Dynamics . . . 337
4.5.5 Attitude Determination and Control . . . 339
4.5.6 Attitude Sensors . . . 341
4.5.7 Actuators for Attitude Control . . . 351
4.5.8 Verification of the Attitude Control System . . . 357
Bibliography . . . 361

4.6 Data Management . . . 361
4.6.1 Data and Information Management On-board . . . 361
4.6.2 On-board Computer . . . 368
4.6.3 Software . . . 371
4.6.4 Dependability . . . 375
Bibliography . . . 379

4.7 Communication . . . 380
4.7.1 Introduction . . . 380
4.7.2 Radio Spectrum . . . 381
4.7.3 Channel Capacity . . . 382
4.7.4 Antennas . . . 383
4.7.5 Thermal Noise . . . 385
4.7.6 Modulation . . . 386
4.7.7 Pulse Code Modulation (PCM) . . . 391
4.7.8 Packet Telemetry . . . 393
4.7.9 Code Division Multiple Access (CDMA) . . . 393
4.7.10 Coupling Networks . . . 394
4.7.11 Transmit and Receive System of the BIRD Satellite . . . 395
Bibliography . . . 398

5 Aspects of Human Space Flight . . . 401

5.1 Humans in Space . . . 402
5.1.1 The Crew of the International Space Station . . . 402
5.1.2 Astronaut Training . . . 406
5.1.3 Infrastructure for Astronaut Training at the EAC . . . 412
Bibliography . . . 418

5.2 Life Support Systems . . . 418
5.2.1 Functions of a Life Support System . . . 418
5.2.2 Metabolic Balances . . . 434
5.2.3 The ISS Life Support System . . . 435
5.2.4 Biological Life Support Systems . . . 437
Bibliography . . . 442

5.3 Rendezvous and Docking . . . 443
5.3.1 Introduction . . . 443
5.3.2 The RVD Mission . . . 443

5.3.3 Basics of Relative Motion . . . 446
5.3.4 The Safety Requirements . . . 449
5.3.5 The ATV RVD System . . . 450
5.3.6 Verification and Test . . . 457
Bibliography . . . 457

6 Mission Operations . . . 459

6.1 Spacecraft Operations . . . 459
6.1.1 General Concepts and Principles . . . 460
6.1.2 Mission Types . . . 461
6.1.3 Premission Support and Implementation Activities . . . 462
6.1.4 Mission Phases . . . 467
6.1.5 Mission Operations Tasks . . . 470

6.2 Control Center . . . 473
6.2.1 Control Rooms . . . 473
6.2.2 Hardware Components . . . 475
6.2.3 Software Components . . . 477
6.2.4 Communications . . . 483

6.3 The Network of Ground Stations . . . 484
6.3.1 The Functions of a Ground Station . . . 484
6.3.2 Site Selection of a Ground Station . . . 487
6.3.3 Subsystems of Ground Stations for Orbiting Satellites . . . 488
6.3.4 Link Design Aspects . . . 494
6.3.5 Ground Station Operation . . . 504

6.4 Operations for Human Space Flight . . . 505
6.4.1 Preparation . . . 506
6.4.2 System Operations of an ISS Module . . . 511
6.4.3 Coordination of ISS Payload Operations . . . 512
6.4.4 The ISS Communication Infrastructure . . . 514

7 Utilization of Space . . . 519

7.1 Earth Observation . . . 523
7.1.1 Categories of Earth Observation Applications . . . 523
7.1.2 Elements of Earth Observation Missions . . . 529
7.1.3 Utilization Programs and Important Earth Observation Missions . . . 536
Bibliography . . . 538

7.2 Communications . . . 538
7.2.1 The Beginning – Sputnik . . . 539
7.2.2 Satellite Communication Services . . . 540
7.2.3 Low-Orbit Spacecraft . . . 541
7.2.4 Satellites in Medium-Altitude Orbits . . . 543
7.2.5 Satellites in High Orbits . . . 544

7.2.6 Satellites in Highly Inclined Orbits . . . 544
7.2.7 Satellites in Inclined, Geostationary Orbits . . . 545
7.2.8 Polar Orbiting Satellites . . . 546
7.2.9 Platforms in the Stratosphere . . . 546
7.2.10 Telecommunication Services: Little–Big–Mega . . . 547
7.2.11 Transponders . . . 548
7.2.12 Transmission Techniques . . . 551
7.2.13 Multiple Access Techniques . . . 551
7.2.14 Frequency Ranges in the Electromagnetic Spectrum . . . 551
7.2.15 Disposal of Satellites . . . 552
7.2.16 Outlook . . . 552
Bibliography . . . 553

7.3 Navigation . . . 553
7.3.1 Basic Principles of Satellite Navigation . . . 553
7.3.2 Satellite Navigation Systems . . . 555
7.3.3 Space Segment . . . 557
7.3.4 Ground Segment . . . 562
7.3.5 Navigation Signals and Services . . . 568
7.3.6 Receiver . . . 572
7.3.7 Accuracy and Error Factors . . . 573
Bibliography . . . 575

7.4 Space Astronomy and Planetary Missions . . . 575
7.4.1 Astronomy Missions . . . 575
7.4.2 Moon Missions . . . 578
7.4.3 Planetary Missions . . . 581
7.4.4 Mission Analysis of Interplanetary Space Probes . . . 587
7.4.5 Key Technologies for Planetary Missions . . . 588
Bibliography . . . 591

7.5 Materials Science . . . 591
7.5.1 Microgravity . . . 592
7.5.2 Critical Phenomena . . . 595
7.5.3 Fluid Physics . . . 596
7.5.4 Solidification . . . 598
7.5.5 Thermophysics . . . 601
7.5.6 Payloads . . . 603
Bibliography . . . 606

7.6 Space Medicine and Biology . . . 606
7.6.1 Medicine in Space . . . 606
7.6.2 Mission Scenarios . . . 607
7.6.3 Experience to Date . . . 609
7.6.4 Environmental Parameters . . . 610
7.6.5 Medical Physiological Problems Arising from Residence in Space . . . 610
7.6.6 Psycho-Physiological Problems Arising from Residence in Space . . . 615

7.6.7 Corrective Measures . . . 617
7.6.8 Outlook . . . 620
Bibliography . . . 620

7.7 New Technologies and Robotics . . . 621
7.7.1 Space Robotics . . . 622
7.7.2 On-Orbit Servicing . . . 637
Bibliography . . . 645

8 Spacecraft Design Process . . . 647

8.1 Mission Concept and Architecture . . . 647
8.1.1 Elements of a Space Mission . . . 647
8.1.2 Segments of a Space Mission . . . 652
8.1.3 The Mission Architecture . . . 653
8.1.4 Development of a Mission Concept and a Mission Architecture . . . 653
Bibliography . . . 657

8.2 Systems Design and Integration . . . 657
8.2.1 Systems Design of a Space System . . . 657
8.2.2 System Integration . . . 662
8.2.3 System Verification . . . 664
Bibliography . . . 667

8.3 Environmental Tests and Basic Concepts . . . 668
8.3.1 Principles and Significance of Environmental Tests . . . 668
8.3.2 Verification Planning and Cost Factors . . . 669
8.3.3 Mechanical Tests . . . 671
8.3.4 Space Simulation Tests . . . 682
8.3.5 EMC and Magnetics . . . 688
8.3.6 Special Environmental and Functional Tests . . . 692
8.3.7 Future Developments . . . 695
Bibliography . . . 696

8.4 System Design Example: CubeSat . . . 696
8.4.1 Introduction . . . 696
8.4.2 Mission Concepts and Scenarios . . . 697
8.4.3 Requirements . . . 698
8.4.4 System Design and Subsystems . . . 699
8.4.5 Model Philosophy . . . 702
8.4.6 Assembly, Integration and Test . . . 704
8.4.7 Operations Aspects and Ground Segment . . . 705
Bibliography . . . 706

8.5 Exemplary System Design of a Microsatellite Mission . . . 706
8.5.1 Microsatellite Design Philosophy . . . 706
8.5.2 Design and Mission Elements of BIRD . . . 707

8.5.3 System Integration and Verification . . . 714
Bibliography . . . 717

8.6 Galileo Satellites . . . 718
8.6.l System Requirements . . . 718
8.6.2 Design Driver and Design Process . . . 719
8.6.3 Platform and Subsystems . . . 721
8.6.4 The Galileo Payload . . . 726
8.6.5 Launcher Interfaces . . . 730
8.6.6 Satellite Assembly, Integration and Testing (AIT) . . . 730
Bibliography . . . 737

9 Management of Space Projects . . . 739
Bibliography . . . 740

9.1 Management of Space Projects . . . 740
9.1.1 Project Management . . . 740
9.1.2 Space Project Characteristics . . . 744
9.1.3 Disciplines of Project Management . . . 752
9.1.4 Tools of Project Management . . . 758
9.1.5 Project Management Documentation . . . 763
9.1.6 Customer–Contractor Relations . . . 765
9.1.7 Conclusions . . . 768
Bibliography . . . 768

9.2 Quality Management . . . 768
9.2.1 Terms . . . 769
9.2.2 Requirements and Premises . . . 770
9.2.3 The Product Main Processes . . . 772
9.2.4 The Organization of Quality Management . . . 775
9.2.5 Product Assurance (PA) . . . 776
9.2.6 Product Assurance for a Project . . . 781
9.2.7 Risk Management . . . 790
9.2.8 Configuration Management . . . 790
9.2.9 Failure and Change Management . . . 791
9.2.10 Requirement Verification . . . 791
9.2.11 Lessons Learned . . . 792
9.2.12 Summary . . . 793
Bibliography . . . 793

9.3 Cost Management . . . 793
9.3.1 Introduction . . . 793
9.3.2 Objective . . . 794
9.3.3 Cost Management Process . . . 795
9.3.4 Tasks of Cost Management . . . 797
9.3.5 Close-out . . . 807
9.3.6 Outlook . . . 807
Bibliography . . . 808

9.4 Legal Aspects of Space Activities 808
9.4.1 Basic Rules of Space Law 808
9.4.2 Legal Conception of a Space Mission 813
9.4.3 Space Activities in the Framework of the EU and ESA 832
Bibliography 834

Acronyms and Abbreviations 837

Symbol List 848

Index 852



Astronaut Thomas Reiter works with equipment for a science experiment on board the International Space Station during his second space mission (Source: ESA)

Foreword

Spaceflight - it is hard to imagine another field of engineering science operating so closely at the frontier of our technological capabilities, while at the same time drawing on the resources of so many diverse scientific disciplines. Developing and operating space systems means achieving the lowest possible structural mass at the highest levels of efficiency and reliability under extreme environmental conditions of temperature, radiation and vacuum.

Accordingly, the ability to develop and operate space systems is not only an indicator of the technical, scientific and industrial capacity of an individual country or an alliance of nations, but also a factor which significantly influences its economic competitiveness. Space activities are a powerful propellant for technical innovation.

Today satellite-based communication, navigation and weather forecasting are an integral part of daily life. Global monitoring of our world has become a necessity for studying climate development. Reconnaissance satellites equipped with a wide variety of instruments are as essential for disaster management as they are for establishing an adequate security policy.

A multitude of probes are currently on a journey through outer space and will provide us with new insights in the area of physics and planetary evolution in our solar system. Space activities have already become a much larger part of our lives than many of us realize.

The development challenges described above are especially applicable for the development human space flight systems. There is hardly a terrestrial transport system, no research station, however remote, in which the lives of the people working there are as dependent on the proper functioning and precise interactions of so many subsystems as in a rocket or space station.

From July to December 2006, for almost six months, I had the opportunity to live and work together with an American and a Russian colleague on board the International Space Station. Although our primary task was scientific research in a variety of disciplines, the effort required of us to operate such a station was relatively high. Meanwhile the European Columbus Laboratory went into operation and when the crew is expanded to six people, as planned for mid 2009, it will be possible to considerably increase the capacity to conduct experiments on board the ISS. As far as the design of future human space flight systems is concerned, it will not only be possible to learn a great deal while living and working in the ISS; the station can also be used as a test environment for new technologies or improving existing ones. There is for example the need to improve methods for the analysis, diagnosis, maintenance and repair of on-board systems, the further development of regenerative life support systems, and the testing of innovative propulsion systems.

The major space agencies in West and East are taking the first steps to prepare for a return to the Moon. The question arises of what role Europe will play in these remarkable and inspiring projects. Drawing on its universities, research institutions and industrial capacity, Europe represented by the European Space Agency (ESA) could make a significant contribution to future human spaceflight.

Viewed against the background of Europe's historical development with its wealth of explorers, researchers and scientists, I definitely regard space activities with all their diverse technical, scientific and industrial aspects as a cultural responsibility as well.

I wish all readers of this handbook a pleasurable and rewarding experience, that pinch of intuition which is so often necessary when crossing borders and gaining new insights, and, not least, a very generous portion of curiosity and enthusiasm for their space activities.



Preface

Dear reader,

The first German edition of the *Handbook of Space Technology* was published in 1988, the second 11 years later. Over the past 20 years there have been profound changes in the engineering, materials, processes and even the politics associated with space technologies and their application. As a consequence, the third edition of the handbook, which you find here in its English version, has been entirely rewritten. The editors proceeded in the same way in which they would approach the task of developing a space system: a number of components must be integrated into a system, which is then expected to fulfill its purpose.

So just as with any proper space mission, we begin with the overall objective, the *Mission Statement*:

The *Handbook of Space Technology* is intended to acquaint students, engineers and physicists, as well as readers with a serious interest in space activities, with the design, construction and operation of a space system. This book also leads the way to a deeper understanding of the corresponding specialized areas.

From this *Mission Statement* requirements can be derived. The *Handbook of Space Technology* should:

- be readable as a whole, although it may require some perseverance while also providing adequate insights and information on selected topics;
- present an overview of a space system in its entirety;
- explain the underlying procedures for design, construction and operation;
- contain references connecting separate chapters as an aid to increasing the reader's understanding.

The editors have deliberated on how these requirements could best be fulfilled. They decided in light of their own limitations not to write the book exclusively by themselves. After all, a space system is not built by the system engineer alone; the approach is rather to connect a great variety of components and parts in a meaningful way. Accordingly, we have asked numerous specialists to write various chapters, and in some cases sections within other chapters. These sections contain the expertise of each of these specialists, but also fit into the general concept of this book, fulfilling, as we hope, the *Mission Statement*.

When integrating the various contributions the editors had to tread a narrow path. On the one hand, the book was supposed to be readable as a smooth, unified whole. On the other hand, it was to reflect each author's unique approach to his or her field. In addition, the use of formulas and the significance of diagrams and figures vary in the individual sections and chapters reflecting the characteristic style of the subdisciplines. The character of the individual chapters can serve to assist students in determining their own particular area of interest, and thus provide some orientation for their studies.

The editors would like to use this opportunity to express their heartfelt gratitude to all the authors for their contributions and their patience throughout the editorial process. The list of authors on the following pages links them to individual chapters making evident how much expertise has been involved, and shows to whom we owe our gratitude. It is also intended to give the reader an indication of which experts in Europe can be consulted for further technical information.

The English version is based on the third edition of the German *Handbuch der Raumfahrttechnik* published by Carl Hanser Verlag in 2007. In the translation the book has been brought up to date (2008).

The editors wish to thank the publishers, especially Mr Jochen Horn, for their support. They also thank Mrs Monika Ebke (DLR, Oberpfaffenhofen) for her superb coordination of editors, authors and Carl Hanser Verlag. Last but not least, the editors thank Mrs Susan Giegerich and Dr.-Ing. Joachim Kehr (DLR, Oberpfaffenhofen) for their valuable contributions to the translation of the book. The editors also thank Ms Nicky Skinner, Miss Beth Dufour and Mr Eric Willner of John Wiley & Sons, Ltd for their substantial effort in preparing the English version.

The concept on which this book was based led to the following divisions. Following an introduction with a historical survey and the characterization of mission types in Chapter 1, Chapter 2 presents the fundamental preconditions and principles underlying a space system. Chapters 3 through 7 characterize the segments of a space system, from the transfer segment (Chapter 3) to the space vehicle (Chapter 4) and the

ground segment (Chapter 6) to user applications and payloads (Chapter 7). Several specific features of human space flight are described in Chapter 5. The concluding chapters survey the tasks of the system engineer (Chapter 8) and space flight project management, and address issues of quality assurance and space law (Chapter 9).

The editors' desire to present not only an overview of space systems technology, but also an introduction to the individual technical disciplines resulted in a work of considerable substance. This calls for some effort and patience on the part of the reader. It required the same from the editors, as we found out.

Nonetheless, we hope that the fascination of space can be felt again and again when working with this text, and that this will facilitate reading.

Wilfried Ley, Klaus Wittmann, Willi Hallmann

The Editors

Prof. Dr.-Ing. Wilfried Ley, Aachen University of Applied Sciences, Germany
Prof. Dr. rer. nat. Klaus Wittmann, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany, and Aachen University of Applied Sciences, Germany
Prof. Dr.-Ing. Willi Hallmann, formerly of Aachen University of Applied Sciences, Germany

The Authors

Dr.-Ing. Sven Abitzsch, European Space Research and Technology Centre (ESTEC), Noordwijk, the Netherlands (Section 9.3)
Dipl.-Ing. Dirk Rüdeger Albat, European Space Agency (ESA), Paris, France (Sections 3.3, 3.4)
Dipl.-Ing. Christian Arbinger, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 4.5)
Dipl.-Ing. Wolfgang Bärwald, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Berlin-Adlershof, Germany (Section 8.5)
Dipl.-Ing. Ralf Baumgartl, Industrieanlagen-Betriebsgesellschaft mbH (IABG), Ottobrunn, Germany (Section 8.3)
Dipl.-Ing. Torsten Bieler, European Space Research and Technology Centre (ESTEC), Noordwijk, the Netherlands (Section 9.3)
Prof. Dr.-Ing. Joachim Block, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Braunschweig, Germany (Section 4.1)
Dr. rer. nat. Hans Bolender, European Space Agency (ESA), European Astronaut Centre (EAC), Cologne, Germany (Section 5.1)
Prof. Dr.-Ing. Klaus Brieß, Technische Universität (TU), Berlin, Germany (Chapter 8, Sections 8.1, 8.2, 8.5)
Dipl.-Ing. Hans Dodel, formerly of Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 7.2)
Prof. Dr. rer. nat. Ivan Egry, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Cologne, Germany (Section 7.5)
Dr. rer. nat. Reinhold Ewald, European Space Agency (ESA), European Astronaut Centre (EAC), Cologne, Germany (Section 5.1)
Prof. Dr. rer. nat. Berndt Feuerbacher, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Bremen, Germany (Chapter 7)
Dr. iur. Michael Gerhard, European Aviation Safety Agency (EASA), Cologne, Germany (Section 9.4)
Dr.-Ing. Anton Grillenbeck, Industrieanlagen-Betriebsgesellschaft mbH (IABG), Ottobrunn, Germany (Sections 3.5, 8.3)
Prof. Dr. med. Dipl.-Geol. Hanns-Christian Gunga, Charité – Universitätsmedizin Berlin, Germany (Section 7.6)
Dr.-Ing. Oskar Josef Haidn, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Lampoldshausen, Germany (Section 3.3)
Prof. Dr.-Ing. Willi Hallmann, formerly of Aachen University of Applied Sciences, Germany (Chapter 1, Section 1.1)
Dr.-Ing. Klaus Hannemann, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Göttingen, Germany (Section 2.3)
Dr. (PhD) Nicolaus Hanowski, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 1.2)
Dipl.-Ing. Martin Häusler, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Weilheim, Germany (Section 6.3)
Dipl.-Ing. Bernward Heese, Astrium GmbH Space Transportation, Bremen, Germany (Section 3.2)
Dipl.-Ing. Christian Henjes, Industrieanlagen-Betriebsgesellschaft mbH (IABG), Ottobrunn, Germany (Sections 2.1, 8.3)
Dipl.-Ing. George Hiendlmeier, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 6.2)

Prof. Dr.-Ing. Gerhard Hirzinger, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 7.7)

Dipl.-Ing. Horst Holsten, formerly of Astrium GmbH Space Transportation, Bremen, Germany (Chapter 3)

Dipl.-Ing. Ulf Hülsenbusch, Industrieanlagen-Betriebsgesellschaft mbH (IABG), Ottobrunn, Germany (Section 8.3)

Dr.-Ing. Clemens Kaiser, Kayser-Threde GmbH, Munich, Germany (Section 7.7)

Prof. Dr.-Ing. Hakan Kayal, University Würzburg, Germany (Section 8.4)

Dr.-Ing. Peter Kern, EADS Astrium Space Transportation, Friedrichshafen, Germany (Section 5.2.4)

Prof. Dr. med. Karl Kirsch, Charité – Universitätsmedizin Berlin, Germany (Section 7.6)

Dipl.-Ing. Joachim Klein, Industrieanlagen-Betriebsgesellschaft mbH (IABG), Ottobrunn, Germany (Section 9.1)

Dr.-Ing. Andreas Kohlhase, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 4.3)

Dipl.-Ing. Johannes Kreuser, CEROBEAR GmbH, Herzogenrath, Germany (Section 4.1)

Dipl.-Ing. Jörg Krüger, Astrium GmbH Space Transportation, Bremen, Germany (Sections 3.2, 3.5)

Dipl.-Ing. Thomas Kuch, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Chapter 6, Sections 6.2, 6.4)

Dr. rer. nat. Holger Kügler, Industrieanlagen-Betriebsgesellschaft mbH (IABG), Ottobrunn, Germany (Sections 2.1, 8.3)

Dipl.-Ing. Hans-Peter Kuhlen, European Aeronautic Defence and Space Company (EADS) Deutschland GmbH, Munich, Germany (Section 8.6)

Dipl.-Ing. Klaus Landzettel, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 7.7)

Dipl.-Ing. Günter Langel, European Aeronautic Defence and Space Company (EADS) Deutschland GmbH, Munich, Germany (Section 3.3)

Dr.-Ing. Jens Laßmann, NGL Prime SpA Turin, Italy (Sections 3.1, 3.2)

Prof. Dr.-Ing. Wilfried Ley, Aachen University of Applied Sciences, Germany (Chapters 4, 9, Sections 2.1, 4.2, 8.3, 8.4)

Dr.-Ing. José M. A. Longo, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Braunschweig, Germany (Section 2.3)

Dr.-Ing. Bernhard Lübke-Ossenbeck, OHB-System AG, Bremen, Germany (Section 4.5)

Dipl.-Ing. Helmut Luttmann, European Aeronautic Defence and Space Company (EADS) Space Transportation GmbH, Bremen, Germany (Chapter 5)

Dr. rer. nat. Manfred Magg, Industrieanlagen-Betriebsgesellschaft mbH (IABG), Ottobrunn, Germany (Section 2.1)

Dipl.-Ing. Jürgen Mathes, OHB-System AG, Bremen, Germany (Section 9.2)

Dr.-Ing. Fritz Merkle, OHB-System AG, Bremen, Germany (Chapter 4)

Dr. rer. nat. habil. Oliver Montenbruck, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 2.2)

Dr.-Ing. Sergio Montenegro, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Bremen, Germany (Section 4.6)

Dr.-Ing. Michael H. Obersteiner, European Aeronautic Defence and Space Company (EADS) Astrium Space Transportation, Les Mureaux, France (Section 3.1)

Dr.-Ing. Willigert Raatschen, European Aeronautic Defence and Space Company (EADS) Astrium Space Transportation, Friedrichshafen, Germany (Section 5.2)

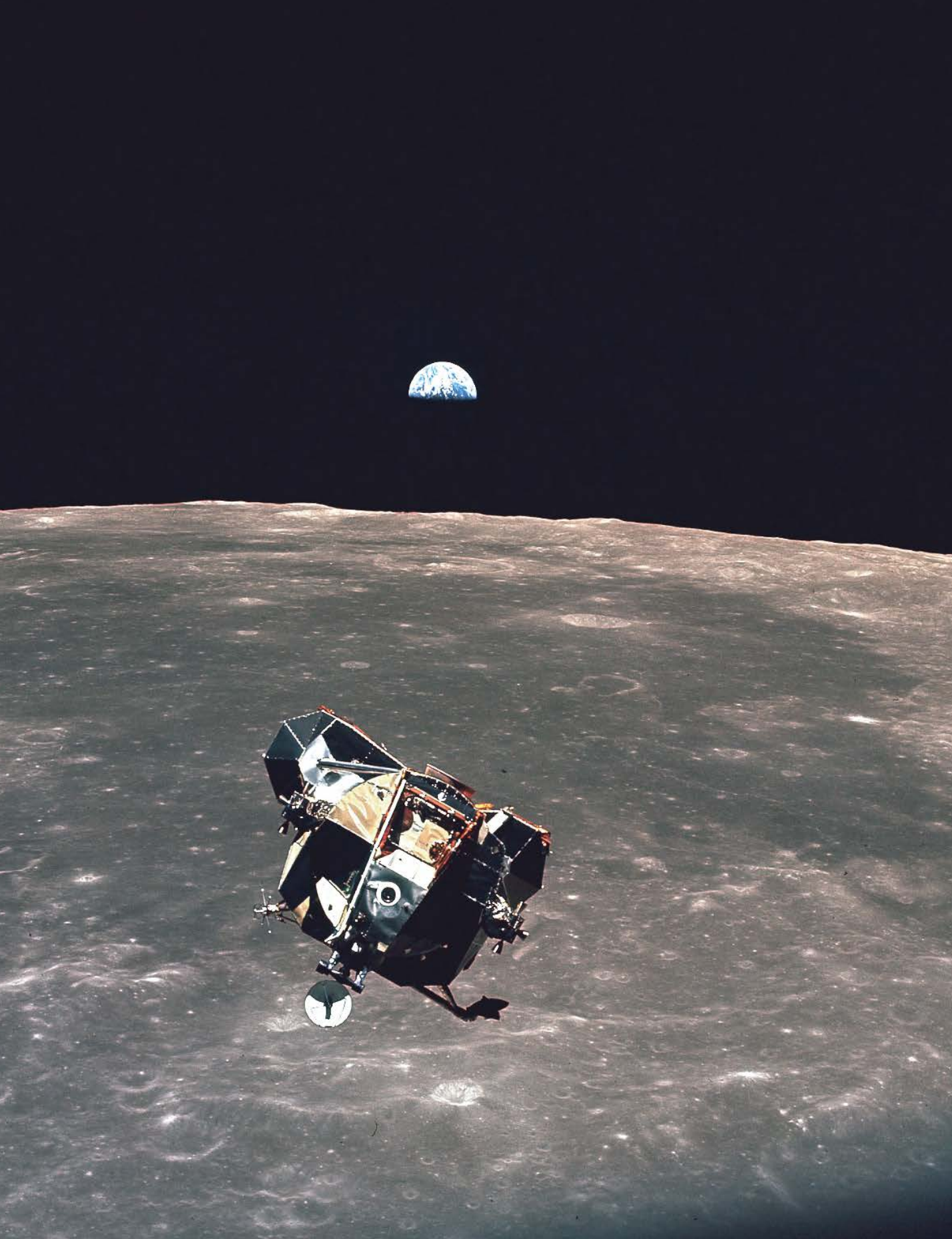
Prof. Dr.-Ing. Hans-Günther Reimerdes, RWTH Aachen University, Germany (Section 2.4)

Dr.-Ing. Klaus-Dieter Reiniger, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 7.1)

Dipl.-Ing. Thomas Reiter, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Cologne, Germany, Member of Executive Board (Foreword)

Dipl.-Ing. Reinhard Röder, European Aeronautic Defence and Space Company (EADS) Astrium Space Transportation, Friedrichshafen, Germany (Section 4.2)

Dr.-Ing. Dieter Sabath, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 6.4)
Dr.-Ing. Stefan Sassen, European Aeronautic Defense and Space Company (EADS) ASTRIUM Services, Munich, Germany (Section 7.3)
Dipl.-Ing. Steffen Scharfenberg, Industrieanlagen-Betriebsgesellschaft mbH (IABG), Ottobrunn, Germany, and ETS BV, Noordwijk, the Netherlands (Section 2.1)
Dr.-Ing. Reinhard Schlitt, OHB-System AG, Bremen, Germany (Section 4.3)
Dipl.-Ing. Hans-Dieter Schmitz, formerly of EADS Astrium Space Transportation, Lampoldshausen, Germany (Section 4.4)
Dipl.-Ing. Alf Schneider, Industrieanlagen-Betriebsgesellschaft mbH (IABG), Ottobrunn, Germany (Section 8.3)
Dipl.-Ing. Artur Scholz, formerly of Aachen University of Applied Sciences, Germany (Section 8.4)
Dipl.-Geophys. Gunter Schreier, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 7.1)
Dipl.-Ing. Josef Sommer, European Aeronautic Defence and Space Company (EADS) Space Transportation GmbH, Bremen, Germany (Section 5.3)
Mathias Steinach, Charité – Universitätsmedizin Berlin, Germany (Section 7.6)
Dipl.-Ing. Peter Turner, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany (Section 4.7)
Dr. rer. nat. Stephan Ulamec, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Cologne, Germany, and Berlin-Adlershof, Germany (Section 7.4)
Dr. rer. nat. Manfred Warhaut, European Space Agency (ESA), European Space Operations Centre (ESOC), Darmstadt, Germany (Section 6.1)
Dipl.-Ing. Klaus Wasserberg, European Space Agency (ESA), European Astronaut Centre (EAC), Cologne, Germany (Section 5.1)
Dr. med. Andreas Werner, Charité – Universitätsmedizin Berlin, Germany (Section 7.6)
Dipl.-Ing. Klaus Wiedemann, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Weilheim, Germany (Section 6.3)
Prof. Dr. rer. nat. Klaus Wittmann, Deutsches Zentrum für Luft- und Raumfahrt eV (DLR), German Aerospace Centre, Oberpfaffenhofen, Germany, and Aachen University of Applied Sciences, Germany (Chapters 1, 2, Section 1.2)



1 Introduction

Klaus Wittmann, Willi Hallmann and Nicolaus Hanowski

The launch of Sputnik 1 in October 1957 marks the beginning of the **space age**. Since 1957 more than 5000 satellites and human spacecraft have entered space and about 850 of them are still operational. The utilization of a spacecraft ends when important subsystems fail or with its controlled or uncontrolled reentry through the Earth's atmosphere. Every year more new spacecraft are launched than old satellites return to Earth. Thus, the number of satellites in space has been continuously growing in the past and is expected to continue to grow in the future. Space flight was initiated by the USA and the former Soviet Union. Since then space projects have been conducted by all major industrialized countries. In addition, a number of developing countries have implemented space programs. On a global scale the USA are still the dominant spacefaring nation according to the number of active spacecraft (see Figure 1.1).

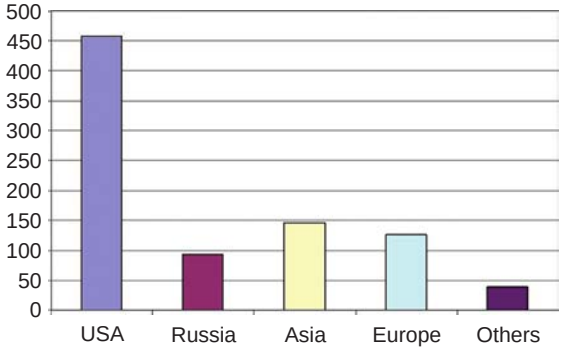


Figure 1.1: *National distribution of operating spacecraft (2007).*

Following the initial period of **non-military space flight**, important **commercial space activities** evolved. These include rocket systems, spacecraft and payloads in the areas of communication, navigation, remote sensing and meteorology.

Space missions are performed not only by single nations, but also by international companies and multinational institutions such as the European Space Agency (ESA).

Cooperation between public entities and commercial companies is gaining importance in space flight projects. These cooperative projects are termed **public–private partnerships** or PPPs. An example of a PPP space project is the German mission TerraSAR-X (Figure 1.2) [1.2] with a high-resolution X-band radar as the main payload.

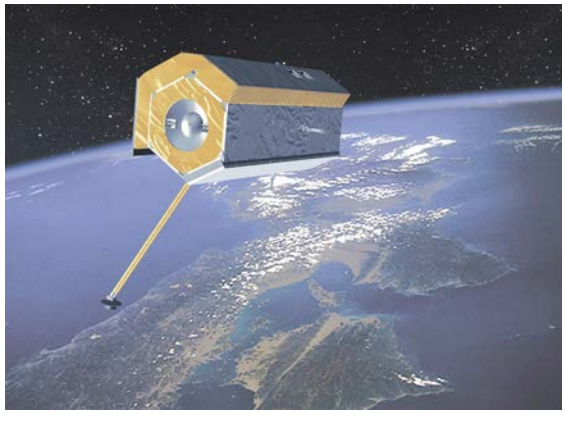


Figure 1.2: *The satellite TerraSAR-X, an example of a PPP mission, is operated by the German Space Operations Centre (GSOC) (Source: ASTRIUM).*

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9

1



Figure 1.3: *The European planetary probe Mars Express which was launched in June 2003 (Source: ESA).*

During the past few decades characteristic **utilization areas** have evolved in space flight. They include the exploration of our planetary system (see Figure 1.3) as well as astronomy and basic research in physics. Observation of the Earth by satellites is carried out for scientific, commercial and military purposes. Communication and navigation missions have gained high commercial value. In technology missions, new systems and components are tested. Human space flight provides a unique environment for research programs including, for instance, experiments in reduced gravity. In addition, exploration of the planetary system by astronauts is in preparation. The importance of satellite missions for military or civil security purposes is recognized by a growing number of nations including the member states of the European Union.

The potential of space missions has been demonstrated over the decade. For the **scientific community** new fundamental knowledge was gained and new fields of research have been opened [1.3]. Space telescopes have improved our knowledge of the Universe because observation became possible in those areas of the electromagnetic spectrum which are not visible from the ground due to atmospheric blocking. By observing the Earth, satellites have also helped to reduce threats and hardship by supporting regional environmental protection and through disaster management. Analysis of effects and identification of the rescue options after natural disasters as well as their prediction increasingly relies on the use of satellites. The potential of space missions for these purposes has not yet been fully exploited. In economic, commercial and transportation areas as well as for individuals, the use of satellites is also gaining importance. The utilization of **satellite communication and navigation** has already become an integral part of our society, growing even more important as the capacity and quality of satellite services continuously improve.

The fascination with space flight also stems from the high visibility of the technical performance needed to develop a space system. Thus, in addition to the direct utilization of space missions, innovation and spin-off products are linked to space flight.

The objective of this book is to provide insight into space systems and the related methods and processes for their development, operation and utilization. Based on practical experience, the state of space flight technology should become apparent. The book also provides an overview of the subsystems typically constituting a space system. In addition the book thoroughly describes the integration of those subsystems into the complete space system. By describing the state of the art, this book also indicates the basis for the development of new concepts and ideas.

Stimulated by the implementation of large space projects such as Galileo, an increase in space activities in Europe can be observed. With new applications and increasing integration of the technical fields involved, a vast development potential for companies has been generated. Academic institutions such as universities are increasingly able to conduct their own satellite missions in order to train their students and to exploit the potential of new technologies. Together with commercial and public space flight activities, this is providing an inspiring and attractive environment for young engineers.

Despite the fascination with space flight, well-trained space technology-oriented engineers are lacking in many European countries. Thus, the education programs in space technology need to be optimized and broadened in order to attract more young people.

A considerable number of European universities and high schools are offering **curricula in space flight technology**. A detailed list would exceed the intended size of this chapter. An excellent entry point for more information is provided by national **organizations** representing the professional community in the aerospace domain. These are, for example:

AAAF Association Aéronautique et Astronautique de France (3AF) (center in Paris, France)
AIAE Asociación de Ingenieros Aeonáuticos de España (center in Madrid, Spain)
AIDAA Associazione Italiana di Aeronautica e Astronautica (center in Rome, Italy)
DGLR Deutsche Gesellschaft für Luft- und Raumfahrt (German Society for Aeronautics and Astronautics; center in Bonn, Germany)
FTF Flygtekniska Föreningen (Swedish Society for Aeronautics and Astronautics; center in Solana, Sweden)
HAES Hellenic Aeronautical Engineers Society (center in Athens, Greece)
NVvL Nederlandse Vereniging voor Luchtvaarttechniek (center in Amsterdam, the Netherlands)
RAeS The Royal Aeronautical Society (center in London, United Kingdom)
SVFW Schweizerische Vereinigung für Flugwissenschaften (Swiss Association of Aeronautical Sciences; center in Emmen, Switzerland)

These organizations have founded a European association, CEAS (Confederation of European Aerospace Sciences), which offers conventions, literature and expert consultancy in the field of space flight and aeronautics.

Public space programs in Europe are initiated and implemented by national space agencies or by ESA.

Examples of national space agencies are:

DLR German Aerospace Centre (center in Cologne, Germany): 29 research institutes and units in 13 locations in Germany including space operations and test sites, for example:
- Space Operations and Astronaut Training:
  – German Space Operations Centre (GSOC) in Oberpfaffenhofen
  – Mobile Rocket Base (MoRaBa) in Oberpfaffenhofen
  – Astronaut Center in Cologne
  – Microgravity User Support Centre (MUSC) in Cologne
- Institute of Space Propulsion Systems in Lampoldshausen

CNES Centre National d'Etudes Spatiales (center in Paris, France)
ASI Agenzia Spaziale Italiana (center in Rome, Italy)
BNSC British National Space Centre (center in London, United Kingdom)
CDTI Centro para el Desarrollo Tecnológico Industrial (center in Madrid, Spain)

Most of these space agencies combine an agency function with research and development functions in order to make new technologies available for their space programs.

ESA, with its head office in Paris, maintains the following research, management and operation facilities:

European Space Research & Technology Centre (ESTEC) in Noordwijk, the Netherlands
European Space Research Institute (ESRIN) in Frascati, Italy
European Space Operations Centre (ESOC) in Darmstadt, Germany
European Astronaut Centre (EAC) in Cologne, Germany
European Space Astronomy Centre (ESAC) in Villafranca, Spain

Further organizations or companies like EUMETSAT (European Organization for the Exploitation of Meteorological Satellites) in Darmstadt, Germany, and EUTELSAT, SES ASTRA, INMARSAT, HISPASAT (Communications), are conducting public and/or commercial space programs. A broad range of companies in the **space industry** is supplying the necessary development potential on system and subsystem scales. Some examples of such European companies are:

EADS (European Aeronautic Defence and Space Company)

1

Headquarters: Schiphol Rijk (the Netherlands)
Within EADS space-related activities are performed by:
- Astrium Satellites
- Astrium Space Transportation
- Astrium Services

Thales Alenia Space
Headquarters: Cannes, Toulouse (France)
Telespazio
Headquarters: Rome (Italy)
OHB (Orbitale Hochtechnologie Bremen)
Bremen (Germany)
Surrey Satellite Technology Limited
Guildford (United Kingdom)
Arianespace
Evry-Courcouronnes (France)

The significance of space technology in research and industrial applications has been recognized by an increasing number of countries all over the world. It is expected that this trend will continue, further increasing the potential of international space programs and providing fascinating new jobs for a worldwide community of scientists and engineers.

Bibliography

[1.1] TerraSAR-X. Das deutsche Radar-Auge im All. DLR-Missionsbroschüre. Deutsches Zentrum für Luft- und Raumfahrt, 2005.
[1.2] Feuerbacher, B., Stoewer, H. *Utilization of Space. Basics, Fields of Usage, Future Developments: Today and Tomorrow.* Heidelberg: Springer Verlag, 2005.
[1.3] 7 Gründe warum Deutschland Raumfahrt braucht. Berlin: Bundesverband der Deutschen Luft- und Raumfahrtindustrie, 2006.
[1.4] Studienangebote Raumfahrt. www.studienwahl.de, 2007.
[1.5] Hallmann, W. *Ingenieure, Wegbereiter der Zukunft.* Düren: Hahne & Schloemer Verlag, 2006.

1.1 Historical Overview

Willi Hallmann

1.1.1 Introduction

The history of space flight is also the history of the rocket. Only a rocket is able to overcome Earth's gravity and travel upward into air-free space. This was not always obvious, as a quote from Max Valier (1895–1930) indicates:

> Just one year ago the problem of rocket propulsion was considered a fairy tale and everyone who fought for it with conviction was derided as a dreamer and laughed at. However, today after the first successful runs of a rocket-propelled vehicle the public is becoming impatient since there is no progress in the advance into space.
>
> *Badische Zeitung,* Karlsruhe, 1929

Hermann Ganswindt (1856–1934), born in Seeburg, East Prussia, may have been one of the first who was convinced about the technical realism of a spacecraft and presented an elaborate construction scheme. He made his first public presentation on May 27, 1891 in the Berlin Philharmonie about his idea of a "world-craft" and explained how space flight might be realized by means of the propulsion principle.

In the twentieth century these visions became reality. Space flight pioneers created the theoretical basis and took the first practical steps. While Konstantin E. Tsiolkovsky (1857–1935) is called the "father of cosmonautics" in Russia, the Americans refer to Robert H. Goddard (1882–1945) as the "father of rocket technology." Hermann Oberth (1894–1989) is considered a "pioneer of space flight" in Europe, while Wernher von Braun (1912–1977) as his ablest student surely did a great deal of the pioneering work as well (see Figure 1.1.1). Not only were technicians excited by the idea of space flight, but also movie makers and artists. Fritz Lang, director of the first space movie *Lady in the moon* (premiered in 1928), introduced the launch countdown, which is still customary today.

Born 1857 in Izhevskoye, Russia, **Konstantin E. Tsiolkovsky** presented his fundamental ideas for space flight in "The conquest of space with propulsion

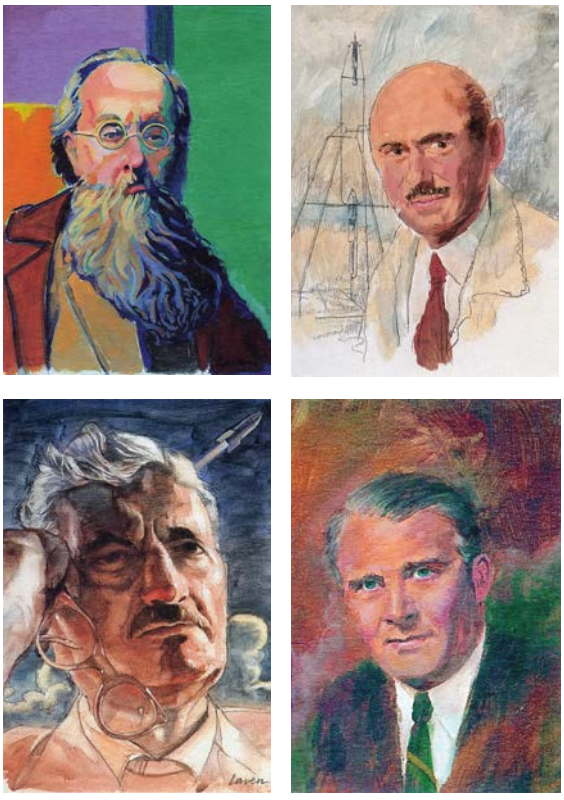


Figure 1.1.1: *Portraits of space flight pioneers: Tsiolkovsky (top left), Goddard (top right), Oberth (bottom left) and von Braun (lower right) (Source [1.1.10]).*

devices"; in 1911 he described an inhabited satellite. He laid the theoretical groundwork of astronautics, and between 1925 and 1932 generated more than 60 papers on that topic.

Born 1882 in Worcester, Massachusetts, **Robert H. Goddard** published a book entitled *About a method to reach greatest altitudes*. In 1926 he launched the world's first successful liquid-fueled rocket (petrol–liquid oxygen). While commercially available rockets were able to produce an emission velocity of 300 m/s, he managed to produce an emission velocity of approximately 2400 m/s with petrol–liquid oxygen.

Hermann Oberth was born 1894 in Hermannstadt, Siebenbürgen. In his book of 1923, *The rocket towards the planet regions*, he described his theory of rocket propulsion in a vacuum.

Technical Data for the A4/V2

Height:	14.03 m
Diameter:	1.68 m
Take-off mass:	12.8 t
Max. velocity:	5.760 km/h
Flight time (fueled):	70 s
Thrust:	approx. 25 t at 2000 m/s exhaust velocity
Max. altitude:	96 km
Range:	330 km

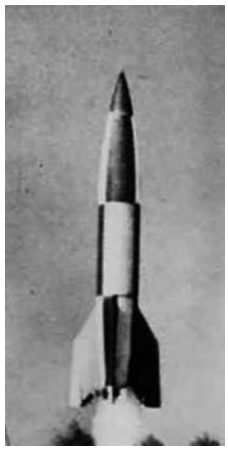


Figure 1.1.2: *Overview of characteristics of A4/V2 rocket used in World War II. (Source: Bundesarchiv).*

The creation of a rocket launch area in Berlin (1930), led by **Rudolf Nebel** (1894–1978), and the use of rockets with liquid fuel were important steps.

The foundation of modern space flight was laid in the years 1935 to 1955. As has been the case several times in the past, technical development was stimulated by war, first by World War II, then by the Cold War.

Military developments in the Soviet Union led to the construction of a two-stage intercontinental vehicle to transport warheads. This development became known as the R7 or "Semyorka." Its further development finally led to the reliable **Soyuz rocket**, today still Russia's only vehicle for human flights. This launcher and the Progress spacecraft trace back to Sergey P. Korolyov.

In May 1945 **Wernher von Braun** and six colleagues were taken into custody by the Americans. In February 1946, 118 engineers and technicians from Germany were working in White Sands, New Mexico. At the beginning von Braun developed the American **medium-range rocket** Hermes C and its derivatives, the Redstone and Jupiter prototypes. The foundation for both Russian as well as American rocket development was originally the German V2 rocket of World War II (Figure 1.1.2). It has been forgotten today that there were considerations in 1950 about using nuclear energy for rocket propulsion [1.1.5], [1.1.6], [1.1.7]. The interested reader is referred to the chapter

1

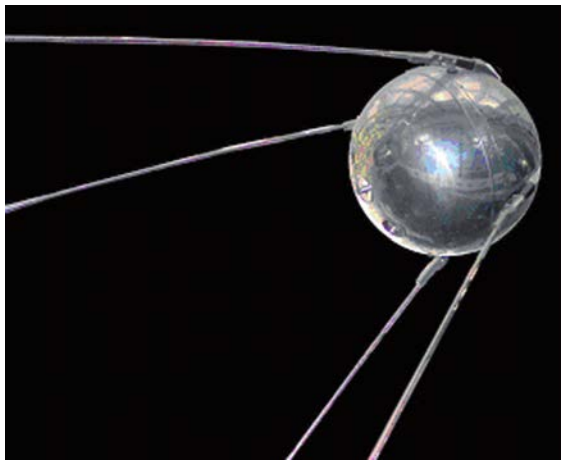


Figure 1.1.3: *With the launch of the first artificial satellite Sputnik 1 on October 4, 1957 the Soviet Union also launched the space age (Source: ESA).*

"Historical overview of the beginnings of space flight," by Ants Kutzer, in the second edition of this handbook [1.1.2] and to [1.1.13], as well as to Ron Miller's interesting publication [1.1.8].

The age of operational space flight began in 1957, when an aluminum sphere with a mass of 83 kg and a diameter of 58 cm excited the world with its signals (Figure 1.1.3). After more than 50 years of experience, space flight is not questioned by anyone.

1.1.2 The Development of Unmanned German and European Space Flight

This historic part of the book has been written from a distinctly German perspective. However, some aspects of the following paragraphs exemplify developmental steps in countries comparable to Germany or even have general implications on how space flight developed.

Initial steps in Germany were the first research projects and experiments for space flight applications starting as early as 1951, when the North German Society for Space Flight was founded. This society launched two test rockets under allied oversight in 1952. In 1954 a German "Aerospace Center" was founded [1.1.3].

In 1962 a modest sum of DM 11 million was allocated for space flight activities. This was not due to the "Sputnik shock" of 1957, but due to the creation of ESRO (European Space Research Organization) and ELDO (European Launcher Development Organization), which Germany joined in 1963. The most important research institutes were in Belgium and the Netherlands.

In Germany a **national space flight program** was set up under the responsibility of an agency originally called the GFW (Society for Space Research). This agency was integrated into the DFVLR (German Aerospace Research and Experiment Institute). At the end of the 1980s this integration was reversed by the founding of DARA (German Agency for Space Affairs) and in 1997 it was reintegrated as part of DLR (German Aerospace Centre). Important satellite missions and human missions (Spacelab, D-1, D-2, etc.) have been conducted as part of the national space program. The technical basis of space flight activities in the German aerospace industry was established in the 1960s, 1970s and 1980s. The resulting knowledge on the component, subsystem and system levels initially led to national satellite missions, contributions to launcher development and human missions. Today DLR institutes, partially in cooperation with industry, develop new sensors, technologies and operation concepts, and are integrated into execution tasks of both the German and international programs.

From 1981 on, the East German **Institute for Cosmos Research** (IFK) that emerged from a number of institutes was also heavily occupied with the development of space flight systems and components. In 1992 the institute was merged with the newly formed DLR site at Berlin-Adlershof.

Independent planning and execution of space flight missions in Germany started in the late 1960s. But especially with respect to launchers, Germany continued to be dependent on the availability of American types. Many of the satellite missions conducted were joint projects in which Germany was able to establish itself as a competent partner.

Important milestones in unmanned space flight were the missions shown in Figures 1.1.4–1.1.17.

As early as 1962 the development of a **launcher system** (EUROPE rocket) commenced on a European level with the objective of creating a European

AZUR
Launch: November 8, 1969
Mass 72 kg, electric power 27 W
First German satellite mission



Figure 1.1.4: *AZUR was intended to explore Earth's radiation belt under the system leadership of the Bölkow GmbH company. The satellite was launched by a US Scout rocket into a polar orbit. A special requirement was that all materials had to be nonmagnetic (Source: DLR).*

capacity to transport a 100 kg payload into a 300 km orbit. The first stage was built by the United Kingdom, the second by France and the third by Germany. Due to several launch failures and for political reasons, the EUROPE rocket program was cancelled in 1972.

In 1975 ELDO and ESRO merged into the newly founded **European Space Agency** (ESA). Since then many highly complex projects have been prepared and conducted under ESA's responsibility. In that program a large number of German contributions were involved. The German budgetary contributions invested in European space flight are distinctly larger than the corresponding national space budgets.

DIAL/WIKA
Launch: March 11, 1970
Mass 63 kg, electric power 10 W
German–French mission



Figure 1.1.5: *In the DIAL/WIKA mission (science capsule) four experiments (among others identifying electron density) were put into an equatorial orbit. The satellite could not be commanded actively. The mission ended after a little more than two months. The mission was launched by a Diamant-B rocket (Source: EADS).*

Finally, European access to space has been assured by the **Ariane** rocket family (Ariane 5 at the moment). And Ariane has now also proven to be economically successful. Important steps were:

1979: The first Ariane rocket (Ariane 1) is successfully launched from the Kourou space center in French Guiana (Figure 1.1.18). The companies Aerospatiale, MATRA, ENRO, MBB and CASA participated significantly in the development and construction of this European satellite launcher.

1984: The 49 m high Ariane 3 is launched for the first time. A version of this rocket without a solid rocket booster became known as Ariane 2.

1990: Aerospatiale receives an order from Arianespace to deliver 50 Ariane 4 rockets.

1

Aeros A and B
Launch: December 16, 1972 and July 16, 1974
Mass 126 kg, electric power 55 W
German atmospheric physics missions



Figure 1.1.6: *The Aeros satellite had a cylindrical structure with a diameter of 0.9 m. One mission objective was to identify the conditions and behavior of the top layers of the atmosphere. In total five experiments were accommodated in each of the satellites. Launch was accomplished by a Scout rocket into a polar orbit (Source: DLR)*

1996: The maiden flight of the new European Ariane 5 takes place but is aborted after 40 seconds because of a software failure.

1997: The 100th flight of an Ariane rocket takes place. In total 134 satellites and 26 piggyback payloads have been put into orbit.

1999: The first commercial use of the Ariane 5 takes place with the launch of the X-ray satellite XMM.

2003: The last launch of Ariane 4 (version 44L) takes place with Intelsat 907 as its payload weighing 4.7 t in total. Up to then, 116 Ariane launchers had put more than 400 t of satellite payload into orbit from Kourou. Three launches failed.

2005: Ariane 5 is launched with the new upper stage ECA and a 10 t payload. This was the 164th Ariane launch.

2006: A new launch with a heavy-duty version ECA takes place. A French and a Japanese satellite are deployed.

Helios A and B
Launch: December 10, 1974 and January 15, 1976
Mass 371 kg, electric power 216/1000 W
German–American solar research mission

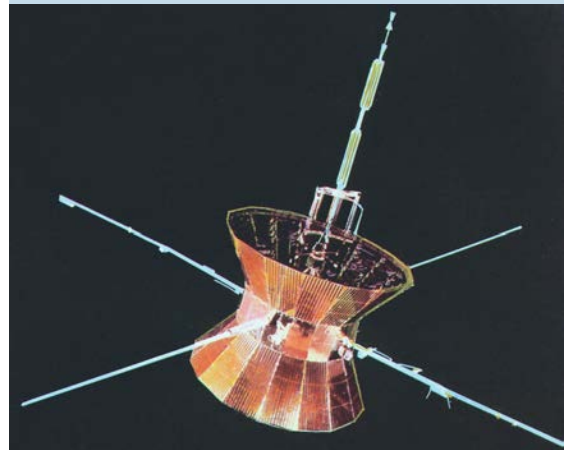


Figure: 1.1.7: *With the solar probe Helios A, which came within 0.3 astronomical units of the Sun, the interplanetary medium in this region was analyzed for the first time. The probe was built in Germany and used by German and American scientists. Launcher: Titan IIIE-Centaur (Source: DLR).*

Significant and ambitious **European space programs** in the areas of astronomy and exploration of the planetary system, Earth observation, navigation and communications are being implemented by ESA. Table 1.1.1 gives an overview of the most important past and current unmanned ESA missions. With the European **Galileo satellite navigation system**, ESA is engaged in a program of considerable magnitude. To implement this navigation system consisting of 30 satellites requires an extensive synthesis of public and industrial competence in Europe. In 2003 the Galileo project was given the go-ahead. In December 2005 the first test satellite for Galileo was launched into orbit

Symphony A and B
Launch: December 19, 1974 and August 27, 1975
Mass approx. 400 kg, electric power 300 W
German–French communications satellite



Figure 1.1.8: *Symphony A, the first German–French experimental communications satellite, was originally planned to transmit the Olympic Games in Munich. The Symphony satellite was three-axis stabilized in geostationary orbit. It was alternately operated by a German and a French control center. Launcher: Thor–Delta (Source: DLR).*

and the second followed in spring 2008. From 2013, after a prior so-called in-orbit validation with four satellites, 30 navigation satellites will provide Europe with an independent global satellite navigation system. It is expected that up to 140 000 new jobs will be created in Europe by Galileo. Ariane 5 will put up to six Galileo satellites at a time into orbit (Figure 1.1.19). Galileo is supposed to ensure Europe's independence but will also be compatible with GPS.

1.1.3 The Development of Human Space Flight in Europe

Human space flight in Europe is built upon the great experience of the Russians and Americans from the 1960s and 1970s. After the first space flight of the Russian Yuri Gagarin (1934–1968) in 1961, efforts by the Soviet Union and the USA were soon directed toward a **human Moon landing**. With resolute preparation, starting with the Mercury program, the USA were able to reach this goal via the Gemini and finally the Apollo programs. In July 1969 Neil Armstrong and Edwin

AMPTE/IRM
Launch: August 16, 1984
Mass approx. 705 kg, electric power 60 W
German–American–British research mission

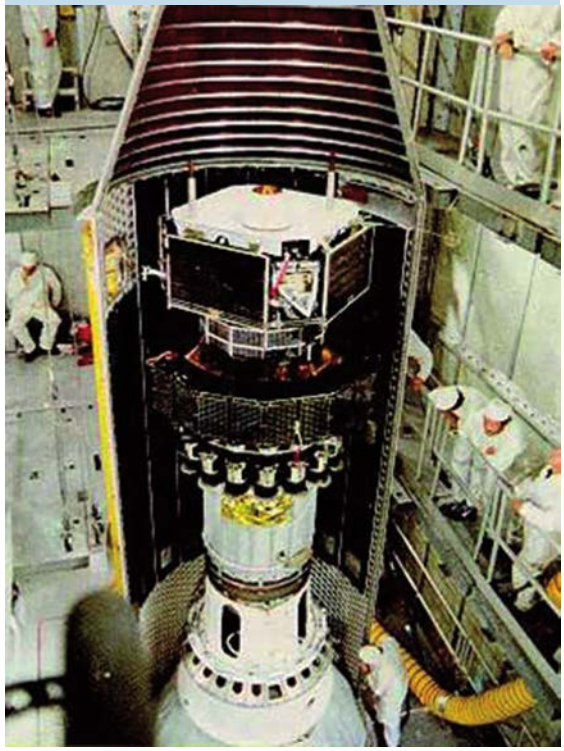


Figure 1.1.9: *AMPTE/IRM (Ion Release Module) was the German contribution to three simultaneously launched satellites for research on the magnetosphere. The satellite deployed barium and lithium and analyzed the behavior of the ion cloud generated. The launch took place with a Delta rocket (Source: NASA).*

Aldrin were the first men to stand on the Moon. The Soviet Union confined itself to the robotic return of lunar samples.

Until the end of 1972, 12 astronauts landed on the Moon as part of the Apollo missions. During this period the two superpowers were already actively initiating the operation of large **space stations** occupied by humans. Important milestones on the way toward a station in orbit for extensive research were:

Salyut 6/7: In April 1971 the Soviet Union put the first space station with two main coupling ports into space. Thus the ISS can be viewed as the grandchild of Salyut 6/7. On August 26, 1978 Sigmund Jähn, a citizen of the German

1

TV-Sat 1 and 2
Launch: November 21, 1987 and August 8, 1989
Mass 2077 kg and 1027 kg respectively, electric power 3 kW
German communications satellite



Figure 1.1.10: *TV-Sat 1 and 2 are direct transmitting satellites whose television and radio signals are strong enough to be received with 50 cm dish antennas. After the deployment of an antenna failed, TV-Sat 1 was placed into a graveyard orbit. Launches took place with Ariane rockets (Source: Aerospatiale).*

DFS-Copernicus 1, 2 and 3
Launch: June 5, 1989, July 24, 1990 and October 12, 1992
Mass 645, 850 and 1400 kg respectively, electric power 1.5 kW
German communications satellite

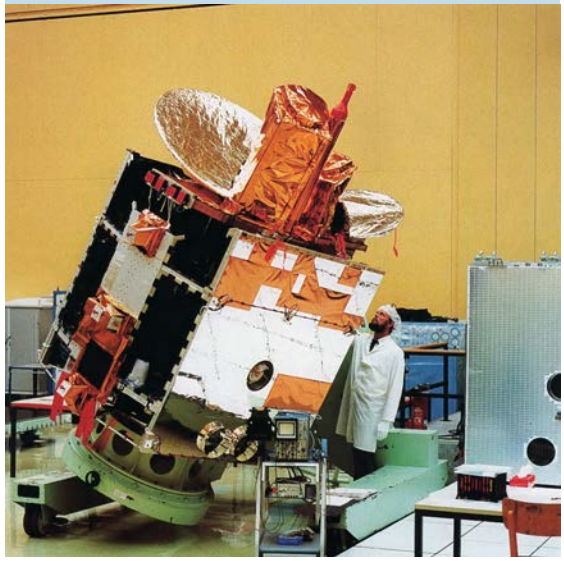


Figure 1.1.11: *The DFS-Copernicus communications satellites were built in Bremen for the German federal postal service. After the launch and early operation phase, conducted by the German Aerospace Centre, the satellites were transferred to Usingen for routine operations. Launchers: Ariane 4 44L and Delta II (Source: MBB/ERNO).*

ROSAT
Launch: June 1, 1990
Mass 2421 kg, electric power 900 W
German–American–British X-ray telescope

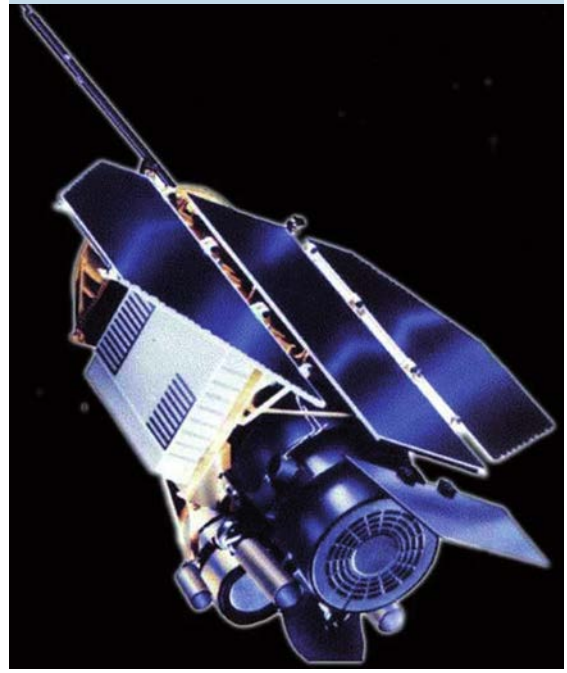


Figure 1.1.12: *On the ROSAT mission a complete survey of the sky for X-ray sources as well as their detailed analysis was conducted. The satellite was three-axis stabilized and operated successfully for almost 10 years. The launch took place with a Delta II rocket (Source: MPG).*

Democratic Republic, was sent aboard Soyuz 31 together with cosmonaut Valery F. Bykovsky to Salyut 6.

Skylab: This US station was placed into an orbit of 432 km altitude and 50° inclination on May 14, 1973. The station comprised a modified third Saturn 5 stage. In the time between May 25, 1973 and February 8, 1974 Skylab was visited by three Apollo command modules with three astronauts each for 28, 59 and 84 days. In July 1979, after more than six years, Skylab reentered the atmosphere and came down over Australia as debris.

MIR: This was a modular space station composed of different station parts which were launched one after another. The assembly started in

EXPRESS
Launch: January 15, 1995
Mass 765 kg
German–Japanese reentry capsule

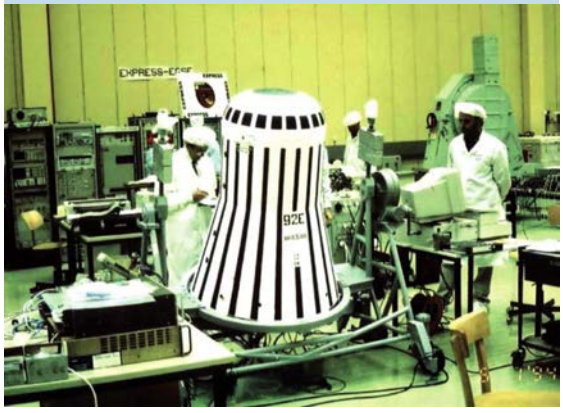


Figure 1.1.13: *Comprising a service and a reentry module, the probe only carried out three Earth orbits due to a launcher failure. Nevertheless, telemetry was received and the reentry vehicle was recovered in Africa. The launch took place with a Japanese M-3SII rocket (Source: DLR).*

EQUATOR-S
Launch: December 2, 1997
Mass 250 kg, electric power 80 W
German research satellite

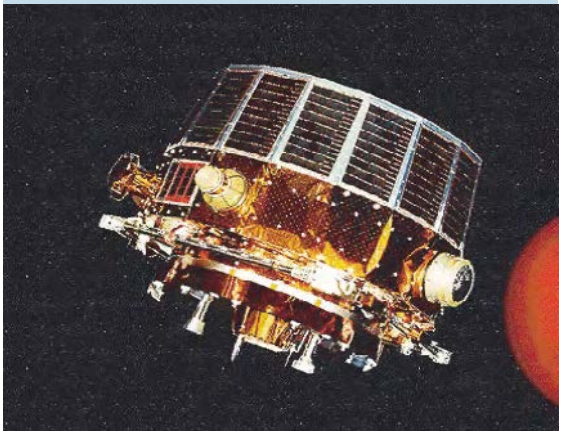


Figure 1.1.14: *The EQUATOR-S satellite was a contribution to the International Solar–Terrestrial-Physics Program (ISTP). It was used to survey plasma, magnetic field and electric field properties at different altitudes. System leadership lay with the Max-Planck Institute for Extraterrestrial Physics. Launcher: Ariane 4 (Source: MPG).*

CHAMP
Launch: July 15, 2000
Mass 522 kg, electric power 140 W
German Earth observation satellite

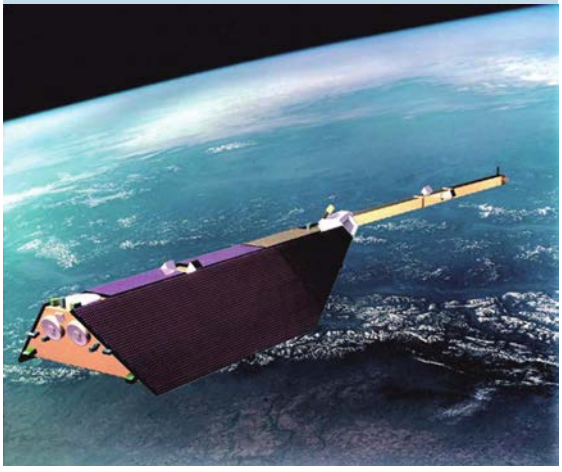


Figure 1.1.15: *With the CHAMP satellite the gravitational field of the Earth, as well as physical and chemical properties of the Earth's atmosphere, are being determined. The payload consists of accelerometers, magnetometers, a GPS receiver, laser retro-reflectors and an ion-drift meter. The launcher rocket was a Cosmos-3M (Source: Astrium/DLR/GFZ).*

February 1986 with the basis module, followed by the Kvant 1 docking module (March 1987), Kvant 2 (November 1989), Kristall (Kvant 3) (May 1990), Spektr (May 1995, docking module for the US Space Shuttle docking in November 1995) and Priroda (April 1996). Except for Kvant 1 the mass of each module was 19 t. In July 1995 the first shuttle docked with the MIR station after the first US astronaut had flown to MIR from Baikonur on a Soyuz spacecraft. The space station was visited by 96 cosmonauts. The longest time on-board was spent by Valeriy V. Polyakov with a total of 679 days, of which 438 days were spent on one mission. The German astronauts Ulf Merbold, Klaus-Dieter Flade, Thomas Reiter and Reinhold Ewald visited the MIR station in the course of the German–Russian missions MIR 92 and MIR 97 and the ESA missions MIR 94 and MIR 95. On April 4, 2000 the last crew were sent to MIR. On March 23, 2001 the 15-year-old station largely burned

1

BIRD
Launch: October 22, 2001
Mass 92 kg, electric power 40 W
German technology satellite



Figure 1.1.16: *With the DLR BIRD satellite numerous new satellite components could be tested. For instance, the infrared cameras provided extraordinary data for detecting and examining fires, volcanic activity and thermal signatures of the Earth's surface. The launcher was an Indian PSLV rocket (Source: DLR).*

GRACE 1 and 2
Launch: March 17, 2002
Mass 490 kg each, electric power 620 W
German–American Earth observation satellites

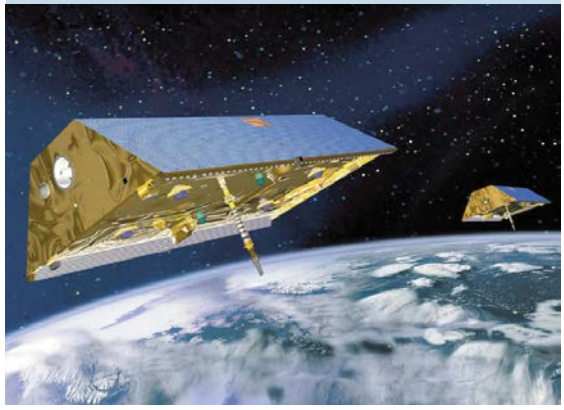


Figure 1.1.17: *Flying with a separation of approx. 200 km the two satellites are used for precise measurements of the Earth's gravitational field. This is achieved by determining variations in the distance between both satellites on a micrometer scale. Launcher: Rokot (Source: Astrium/DLR).*



Figure 1.1.18: *Ariane 1. First successful rocket launch on December 24, 1979 in Kourou (Source: ESA).*

up in Earth's atmosphere after a controlled reentry. However, 19 t of the total 124 t mass crashed unburned into the Pacific Ocean. During its history the MIR space station, which was originally designed for a lifetime of seven years, orbited the Earth 86 325 times at an altitude of 390 km [1.1.12].

With the **Space Shuttle** (first launched in April 1981) a partially reusable and very capable system became available for the USA. The shuttle played an important role in the transport of heavy satellites and laboratory modules and later in the transport of large components to the ISS. High costs and the loss of the *Challenger* (1986) and *Columbia* (2003) shuttles from a fleet comprising a total of five shuttles led to the decision to phase out the shuttle program by 2010 or shortly thereafter.

Table 1.1.1: *Important ESA missions.*

Year	Mission name	Mission application
1968	HEOS 1	Space physics
1975	COS-B	Gamma radiation astronomy
1978	IUE	Ultraviolet space telescope
1978	GEOS 2	Magnetosphere survey
1983	EXOSAT	X-ray astronomy
1985	Giotto	Comet fly-by
1989	Olympus	Experimental communication
1989	Hipparcos	Astrometry
1990	Ulysses	Solar research
1991	ERS-1	Earth observation
1992	EURECA	Experiment platform
1995	ISO	Infrared space telescope
1995	SOHO	Solar research
1997	Huygens	Titan landing probe on Cassini
1999	XMM-Newton	X-ray astronomy
2000	Cluster	Magnetosphere research
2002	INTEGRAL	Gamma radiation astronomy
2002	ENVISAT	Earth observation
2003	SMART-1	Moon exploration
2003	Mars Express	Mars exploration
2004	Rosetta	Comet rendezvous
2005	Venus Express	Venus exploration



Figure 1.1.19: *Artist's impression of the planned Ariane 5 upper stage with eight Galileo satellites prior to separation (Source: ESA).*

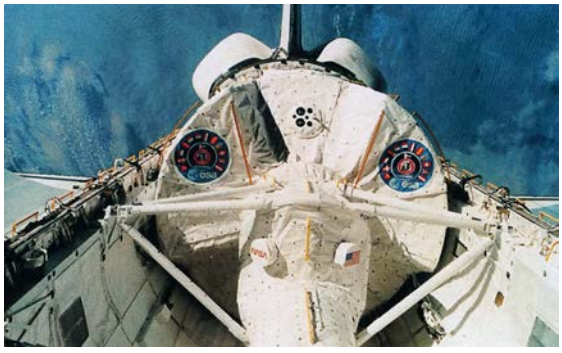


Figure 1.1.20: *Spacelab, built by MBB/ERNO in Bremen, flew onboard the Space Shuttle Columbia for the first time. As the first astronaut of the Federal Republic of Germany, Ulf Merbold was on this European mission which featured 38 experiments. The launch took place in November 1983 (Source: NASA).*

In Europe the **Spacelab**, **SPAS** and **EURECA** platforms were developed by MBB for ESA as contributions to the shuttle program around 1980.

For human space flight ESA facilities work closely with national institutions. German astronauts have been part of important missions, especially for the Spacelab FSLP (1983, Figure 1.1.20), D1 (1985, Figure 1.1.21), D2 (1993) and SRTM (2000) missions which were conducted in cooperation with the USA.

1.1.3.1 The International Space Station

A little more than 40 years after the first space flight, the first of the basic elements for the assembly of the future International Space Station (ISS) (Figure 1.1.22) was launched on November 20, 1998 from Baikonur in Kazakhstan. The corresponding plans date back to the 1980s. At that time the space station was referred to as "Freedom" or "Alpha." The project became a cooperative effort among several nations. In addition to NASA and the Russian space flight agency Roskosmos, Europe is also participating. ESA signed a contract to cooperate in the station's construction in 1998. Moreover, the Canadian and Japanese space agencies have also signed contracts.

As a partner of the USA, Russia, Japan and Canada, Europe operates the **Columbus laboratory module** as

1



Figure 1.1.21: *The Spacelab D1 mission was launched with two German scientists on-board (Ernst Messerschmid and Reinhard Furrer). Mission management and payload operations were a German responsibility. The launch took place on October 30, 1985 with the Space Shuttle Challenger (Source: MBB/ERNO).*



Figure 1.1.22: *Artist's impression of the ISS (Source: ESA).*

part of the ISS and provides an **automated transfer vehicle** (ATV) for supplying the station. In 2006 the Columbus module was handed over to NASA by Germany for integration into the Space Shuttle in order to transport it to the ISS (launched February 2008).

Even in an unfinished state (construction should be finished by 2010) the station has been occupied by astronauts and cosmonauts or tourists from the beginning (see also Table 1.1.2). After completion it will reach a size of approx. 110 m × 90 m × 30 m and will stay in operation at least until 2016. At the moment it is the biggest human-built object in Earth orbit.

Table 1.1.2: *European astronauts and cosmonauts with space experience (as at June 2008).*

Name	Country
Aleksandar Panayotov Alexsandrov	Bulgaria
Anatoly Artsebarsky	Ukraine
Patrick Baudry	France
Ivan Bella	Slovakia
Maurizio Cheli	Italy
Jean-Loup Chrétien	France
Jean-François Clervoy	France
Frank de Winne	Belgium
Pedro Duque	Spain
Reinhold Ewald	Germany
Léopold Eyharts	France
Bertalan Farkas	Hungary
Jean-Jacques Favier	France
Klaus-Dietrich Flade	Germany
Dirk Frimout	Belgium
Christer Fuglesang	Sweden
Reinhard Furrer	Germany
Umberto Guidoni	Italy
Claudie Haigneré	France
Jean-Pierre Haigneré	France
Miroslaw Hermaszewski	Poland
Georgi Ivanov	Bulgaria
Sigmund Jähn	Germany (GDR and FRG)
Leonid Kadenyuk	Ukraine
André Kuipers	Netherlands
Franco Malerba	Italy
Ulf Merbold	Germany
Ernst Messerschmid	Germany
Paolo Nespoli	Italy
Claude Nicollier	Switzerland
Wubbo Ockels	Netherlands
Philippe Perrin	France
Dumitru Prunariu	Romania
Thomas Reiter	Germany
Vladimir Remek	Czechoslovakia/Czech Republic
Hans Schlegel	Germany
Helen Sharman	United Kingdom
Anatoly Solovyev	Lithuania
Gerhard Thiele	Germany
Michel Tognini	France
Franz Viehböck	Austria
Roberto Vittori	Italy
Ulrich Walter	Germany

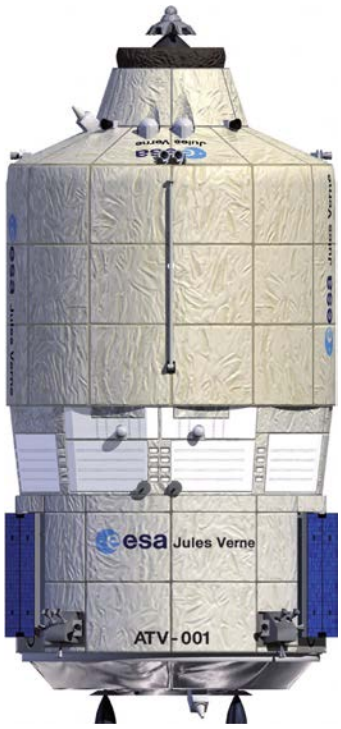


Figure 1.1.23: *The ATV for supplying the ISS (Source: ESA).*

The station is circling at an altitude of 350 km with an inclination of 51.6°. At the end of 2006 astronaut Thomas Reiter completed a long-term stay on the ISS which had begun in July 2006 in the course of the Astrolab mission.

Beginning in 2008 Europe is also contributing to the supply of the space station. This is being accomplished by the ATV (Figure 1.1.23), according to the same principle as for the Russian Progress space transporter. In summer 2004 production of six such ESA transporters was initiated under a contract with EADS Space Transportation. The contract lasts until 2013. The ATV comprises three main elements: a propulsion system, a control unit with an on-board computer, and the payload. Its task is to keep the ISS alive and to supply materials (food/water, oxygen, fuel, experimental equipment, etc.). Its technical data is as follows: overall length, 10.3 m; diameter, 4.48 m; max. take-off mass, 20.75 t; payload, 7.6 t; mission duration, max. six months docked to ISS; power supply, four solar panels and eight rechargeable batteries.



Figure 1.1.24: *The Columbus module attached to the ISS (Source: ESA).*

The **Columbus program** was initiated in 1986 at an EU Council of Ministers conference to be implemented by ESA in addition to the Hermes and Ariane 5 programs. Columbus (Figure 1.1.24) was originally meant to be a laboratory docked to the US Space Station Freedom (SSF) or a free-flying device. The name Columbus was chosen because the discovery of America by Columbus had its 500th anniversary in 1992. Thus hope was expressed that Columbus would dock with the SSF in 1992. However, Columbus was initially also intended to be able to be launched by an Ariane 5. The launch of Columbus and docking with the ISS became a reality in 2008. Operation of the Columbus module is conducted by the Columbus Control Centre at DLR/GSOC Oberpfaffenhofen.

On a personal note, it is unrealistic to illustrate the history of space flight over the last 50 years within 10 pages. Everything stated above has been chosen subjectively and must therefore be incomplete. During preparation, the journals *SGLR-Luft- und Raumfahrt* and *Planet Aerospace* as well as [1.1.14] were of great help.

Bibliography

[1.1.1] Puttkamer, J. v. *Von Apollo zur ISS.* Munich: Herbig Verlag, 2001.

[1.1.2] Hallmann, W., Ley, W. *Handbuch Raumfahrttechnik*, 2. Auflage. Munich: Carl Hanser Verlag, 1999.

[1.1.3] Krieger, W. *Technologiepolitik der Bundesrepublik Deutschland (1949–1990)*, Band IX, S. 242. Düsseldorf: VDI Verlag, 1992.

1

[1.1.4] Hornschild, K., Neckermann, G. *Die deutsche Luft- und Raumfahrtindustrie, Stand und Perspektiven.* Frankfurt a.M.: Campus Verlag, 1988.
[1.1.5] Reichel, R.H. Die heutigen Grenzen des Raketenantriebes und ihre Bedeutung für den Raumfahrtgedanken. *VDI-Z*, **92** (32), 1950.
[1.1.6] Reichel, R.H. Raketenantriebe. *VDI-Z*, **102** (12), 1960.
[1.1.7] Micheley, W. Bericht über den IX Internationalen Astronautischen Kongress 1958 in Amsterdam. *VDI-Z*, **100** (36), 1958.
[1.1.8] Miller, R. *The Dream Machines.* Molabor, FL: Krieger, 1993.
[1.1.9] *Zeit im Flug: Eine Chronologie der EADS.* Hamburg: EADS Edition, 2003.
[1.1.10] Gierson, R. *et al.* DESK CALENDAR 1988, General Dynamics, Space System Division, 1988.
[1.1.11] Messerschmid, E., Bertrand, R. *Space Station Systems and Utilization.* Berlin: Springer Verlag, 1999.
[1.1.12] Gilbert, L., Rebrow, M. *Das Thomas Reiter Kosmosbuch.* Klitzschen: Elbe-Dnjepr Verlag, 1996.
[1.1.13] Engelhardt, W. *Enzyklopädie der Raumfahrt.* Frankfurt a.M.: Harry Deutsch Verlag, 2001.
[1.1.14] Reinke, N. *Geschichte der deutschen Raumfahrtpolitik.* Munich: Oldenbourg Verlag, 2004.







Figure 1.2.1: *The three segments comprising a space system: the space segment with the space vehicle (top), the transfer segment with the launcher (center) and the ground segment with control center and ground station (bottom) (Source: ESA/DLR).*

1.2 Space Missions

Klaus Wittmann and Nicolaus Hanowski

1.2.1 Space System Segments

A typical space flight system comprises three system segments, which are coordinated according to the mission objectives (Figure 1.2.1). The design of the system segments and consideration of their mutual dependencies is the central challenge for successfully preparing and conducting space flight missions.

The **space segment** comprises the spacecraft and its payload in orbit. The **transfer segment** provides the transport of the spacecraft and its payload into space by a launcher (typically a rocket). In order to control and monitor the spacecraft and its payload as well as to distribute and process the payload data, a **ground segment** is required. The design of ground and transfer segments and the costs connected with their realization are mainly influenced by the physical parameters of the spacecraft and the payload. In turn, these depend essentially on the mission objective and the mission duration. The three system segments can be split up further into so-called system elements (Figure 1.2.2).

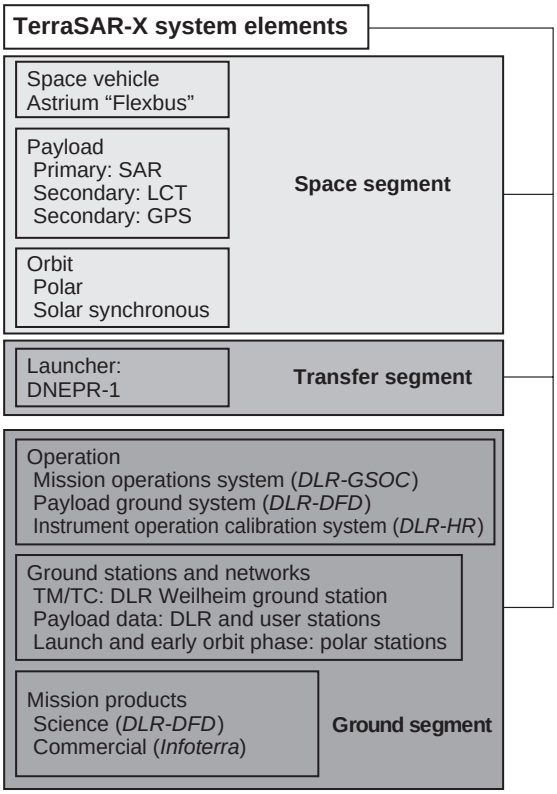


Figure 1.2.2: *Organization of a space flight system in system elements, exemplified by the German radar remote sensing satellite TerraSAR-X (SAR = Synthetic Aperture Radar, LCT = Laser Communication Terminal, GPS = Global Positioning System).*

1.2.1.1 The Space Segment

System Element: Payload

As the central application element the **payload** is at the heart of a space flight mission. Successful payload operations open the door to mission success or put it in doubt, even if all other subsystems of a spacecraft work flawlessly. The payload's proximity to the application and therefore to the actual motivation for the mission justify an extraordinary position for it within the whole system design process (Table 1.2.1).

The payload with its characteristic parameters of mass, geometry, power and communication requirements determines the properties of the carrying **satellite platform**, which is often referred to as the satellite bus. In human space flight there is the additional task of providing life support for the crew.

System Element: Orbit

This system element is essential for conceptualizing a space flight system. The orbit of a spacecraft is defined by the mission objective. More than 95% of all space mission orbits are orbits around the Earth. Low Earth orbits between 300 and 1500 km are often used, for example, for **Earth observation satellites** and human **space flight**, and the so-called geostationary orbit at approx. 36 000 km altitude for **communications satellites** (Figure 1.2.3). Orbits at intermediate altitudes, such as so-called medium Earth orbits (MEOs), are used for instance for **navigation satellites** (GPS, Galileo). The relatively small number of interplanetary missions on which spacecraft are sent beyond an Earth orbit into planetary orbits are often characterized by several years of flight time until the spacecraft reach their target object or orbit (see Table 1.2.2). With the exception of the Apollo missions to the Moon, which were concluded in 1972, planetary missions are still limited to unmanned endeavors.

Unmanned spacecraft flying in Earth orbit are referred to as **satellites**. When flying in orbits beyond Earth's orbit they are termed **space probes**. Spacecraft carrying humans are referred to according to their functions as space shuttles, space ships or space stations. Objects with ballistic trajectories which can reach altitudes in excess of 1000 km are called **suborbital rockets** or sounding rockets. They are not discussed in this book.

System Element: Spacecraft

With ever-expanding areas of application, **spacecraft** have evolved in their development over more than 50 years into a huge variety of types with a wide range of characteristics. However, in order to work properly the spacecraft has to perform an invariable set of functions. The corresponding functional structure of subsystems represents the common basis for design, production and operation of all spacecraft. In particular, the complexity of subsystems has dramatically increased over the decades. Nevertheless, the functional logic of each subsystem as well as aspects of its compatibility have not changed significantly. The following spacecraft subsystems are generally distinguished (see also Figure 1.2.4).

1

Table 1.2.1: *Payload overview with examples of their applications on spacecraft.*

Payload	Application	Characteristic	Mission example
• Cameras (UV/Vis./IR) • Radar	• Earth observation • Weather monitoring • Planetary exploration • Astronomy	• Payload: from global overview down to high resolution of selected spots	• EnMAP • SAR-Lupe • Meteosat • Mars Express • Hubble Space Telescope
• Sensors (nonimaging)	• Earth exploration • Atmospheric research • Planetary exploration	• Great variety of payloads	• CHAMP • GRACE 1 and 2 • ENVISAT
• Experimental components, spacecraft components	• Validating new technology	• Passive, robotic	• BIRD • TerraSAR-X • ROCVISS on ISS
• Repeater/transponder	• Television • Internet • Telephony	• Large satellites, often in geostationary orbit	• EUTELSAT • ASTRA • Iridium
• Signal transmitter • Atomic clock	• Navigation • Positioning	• Typically in medium to high orbits	• GPS • Galileo • Glonass
• Lander • In-situ analysis instruments • Rover	• Analyses of planet surfaces	• Highly complex systems for human spaceflight	• Apollo • Viking • Giotto • Mars Express • Philae/Rosetta

Structure

Numerous immediate characteristics of a spacecraft are determined by its **mechanical structure**, which accommodates all other subsystems. As well as the pure static properties of the structure there are often dynamic aspects, such as deployment, rotation and swing functions, with frequent and considerable effects on other subsystems.

Power Supply

The focus of this subsystem is on assuring an efficient distribution of **electrical energy** within the spacecraft and its components. Power sources can be for instance: solar generators, batteries, fuel cells or so-called radio-isotopic thermoelectric generators (RTGs).

Thermal Subsystem

The temperature of spacecraft components has to be kept within a defined range. Not only are temperature-related tolerances crucial, but so too is the efficiency of components (solar panel, sensors, etc.) under certain temperature conditions. The thermal subsystem provides an optimized equilibrium between heat absorption and dissipation by passive and/or active regulation.

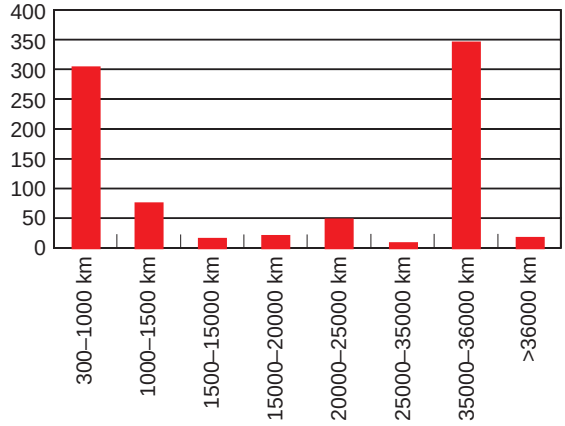


Figure 1.2.3: *Number of operating satellites at various orbital altitudes (average orbit altitude).*

Table 1.2.2: *Overview of orbits for space flight missions.*

Orbit	Application	Characteristic	Mission example
• LEO (Low Earth Orbit)	• Earth observation • Weather monitoring • Technology • Astronomy	• Altitude of 300 up to 1500 km	• CHAMP • SAR-Lupe • BIRD • ROSAT
• MEO (Medium Earth Orbit)	• Communications • Navigation	• Altitude of several thousand km	• Globalstar • GPS • Galileo
• HEO (High Elliptical Orbit)	• Communications • Astronomy	• Altitude of a few hundred up to 100 000 km	• Molniya
• GTO (Geostationary Transfer Orbit)	• Injection orbit for launchers of communications satellites	• Altitude of a few hundred up to 35 786 km	• EUTELSAT • ASTRA
• GEO (Geostationary Orbit)	• Communications	• Altitude of 35 786 km	• EUTELSAT • ASTRA
• Lagrange points	• Astronomy • Fundamental research	• Distance > 1 million km	• SOHO • JWST
• Interplanetary orbits	• Planetary exploration	• Distance up to several billion km	• Mars Express • Rosetta

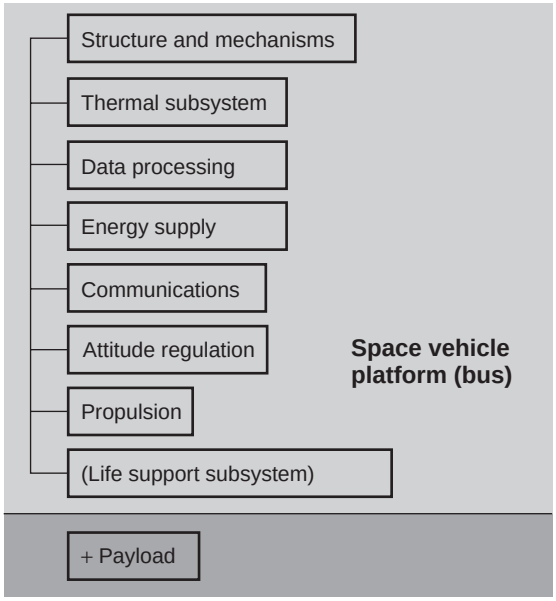


Figure 1.2.4: *Differentiation of spacecraft system elements into subsystems.*

Attitude Control

The attitude control subsystem monitors and controls the **orientation** of the spacecraft in space. In many cases this is the most complex subsystem with a huge number of parameters, sensors, and active and passive control components. Especially, activities such as the use of reaction thrusters or the acceleration of reaction wheels require a good understanding of the orientation and dynamic properties of the spacecraft.

Communications

Central components of this subsystem are transmitters, receivers and antennas. There are different types of data sets to be transmitted to and from Earth or between individual spacecraft: so-called telemetry for spacecraft monitoring, commands for control, and payload data.

Data Processing

In this subsystem the processing and formatting of data generated on the spacecraft are carried out. Central elements are corresponding on-board computers and peripheral equipment. In contrast to the data system hardware, on-board software can still be modified after launch by so-called software uploads.

Propulsion

This subsystem allows the spacecraft's orbit to be changed by firing thrusters. With the application of electric propulsion it has become necessary to master long-lasting propulsion maneuvers. In contrast, typical propulsion phases with chemical thrusters last only minutes or hours.

Life Support System
This system evolved from the special requirements of human space flight. It is limited to this area and guarantees physical integrity and appropriate living conditions for humans in space.

In addition to the function and capability of each subsystem it is very important to consider their compatibility and the properties of the complete system.

System engineers and other **system experts** hold a key position in the design and production phase as well as with regard to operations. Only by systematic design that focuses on the interaction between space and ground segments, corresponding adaptations on the spacecraft system and subsystem levels, and the payload can an optimized space mission be realized.

1.2.1.2 Transfer Segment

Another system segment in space missions involves the launcher that transports the spacecraft into space. Numerous rockets have become available on the commercial market over the years. With the Ariane 5 rocket Europe has a powerful and internationally competitive product at its disposal.

Significant factors in choosing a specific launcher are the orbit to be reached as well as the mass and dimensions of the spacecraft. Due to the high development and modification costs of rockets, the variety of relevant types for a mission profile is often very limited. This also means that the rocket offers fewer variables for mission optimization than do the spacecraft and the ground segment. However, multiple launches of several spacecraft at the same time and additional boost stages on the spacecraft to reach certain orbits offer some additional flexibility. It is also possible for one or more small satellites to be launched "piggyback" together with the main payload.

For spacecraft with a mass under 2 t and LEOs there are a large number of launchers available. At present these are built and launched also by developing countries, such as Brazil or India. On the other hand, there are only a few models available at the upper end of the scale. Satellites with a mass of more than 8 t can only be launched into geostationary orbit by the Ariane 5 ECA as well as by the US Atlas V and Delta IV rockets. For extreme launch masses in the area of 25 t into LEO there is only the Space Shuttle available. Because of the growing use of powerful communications satellites, there is a trend toward rather heavy satellites (Figure 1.2.5).

1.2.1.3 The Ground Segment

In addition to the spacecraft with its payload and orbit, the **ground segment** shapes the space flight mission scenario. Similar to the spacecraft, the ground segment also provides a large set of degrees of freedom in the design with a high potential of optimization with regard to efficiency. In contrast to the spacecraft, considerable changes can still be applied to the ground segment after launch. These changes can sometimes be quite extensive and often decide the success of a mission. Mostly, however, late changes translate into significant additional work and costs.

The ground segment can be divided into two system elements: mission operations and the ground station network.

System Element: Mission Operations
For the most part, mission operations are designed and conducted at a **control center**. With this system the spacecraft is monitored and controlled and the data traffic organized. In addition, the control center contains all necessary interfaces to the spacecraft manufacturer as well as to its users. The control center routes all relevant data to them.

The central part of mission operations is **flight operations**, which are conducted from a control room

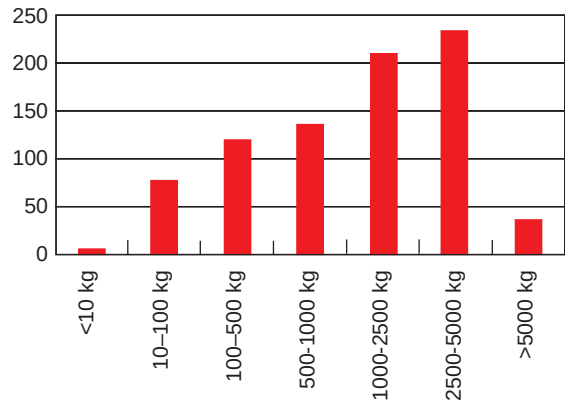


Figure 1.2.5: *Number of operating satellites according to their launch mass (fueled).*

(Figure 1.2.6). Flight operations activities include the management of the actual flight tasks and the maintenance of the spacecraft in all mission phases in order to use it optimally and as long as possible. Operation specialists for the subsystems are organized in a flight team. Under the lead of a responsible system engineer and by means of telemetry from the space segment, their analyses include considering the status and trends in the mission, working through procedures, generating command sequences and conducting flight maneuvers. Mission management is carried out from or in close coordination with flight operations.

Operations activities vary profoundly with the different phases of preparation and execution of a mission. The most intense phase of mission operations is the so-called **launch and early operation phase** (LEOP), in which the spacecraft is activated after injection and its survival secured under the extreme conditions of space with respect to temperature, vacuum, radiation, etc. Compared to the following phases of **in-orbit testing** (IOT) and especially routine operations, this flight phase is by far the most demanding in terms of personnel and resources.

Flight operations use the ground data processing facilities and equipment at the control center. These facilities assure the correct configuration and availability of all **telemetry** data required for operations. Additional contributions to flight operations are made by the flight dynamics group and by mission planning. The former is especially responsible for orbit determination and prediction as well as other aspects of navigation. The latter provides the tools necessary to generate and handle timelines, schedules, etc., and considers user priorities as well as physical and technical constraints. Ground data processing, flight dynamics and mission planning are highly interlinked with flight operations and coordinated by the mission management. In complex missions the aim is to grant all mission participants easy access to relevant data, information and products on electronic platforms.



Figure 1.2.6: *Control room of the Columbus control center as part of the flight operations at DLR in Oberpfaffenhofen. From this room European activities on the ISS are coordinated and monitored (Source: DLR).*

Increasingly, close cooperation between so-called **user ground centers** and mission operations is required. In commercial Earth observation missions these centers take care of tasks such as processing raw user data into finished information products. Data refining, thematic editing and archiving are some of the relevant aspects. Sometimes necessary payload activities, including instrument calibration and configuration changes, are also prepared. Other services include interface adaptations for routine operations.

System Element Ground Station Network

The **antennas** (Figure 1.2.7) located at ground stations are the most visible element of the transmission path from and to the spacecraft. Various frequency bands can be used whose allocation is subjected to international coordination. Ground stations receive the data for spacecraft control from an associated control center. On the other side, status and user data from the spacecraft are routed to the control center from the ground station. In addition, a ground station provides orbit determination (tracking) data.

The mission profile of a ground station is also highly dependent on the mission phase. During mission preparation, compatibility with communication components of the spacecraft and configuration of data interfaces have a high priority. With separation from the launcher and the beginning of the LEOP, detection of spacecraft position as well as frequent and long-lasting contacts with the spacecraft are especially important. Visibility of satellites in LEO typically lasts only a few minutes for one ground station. After that the spacecraft is without contact to the ground for periods from 90 minutes up to several hours, if no other ground stations are employed. In order to assure more frequent or even around-the-clock contact with the spacecraft, especially in the critical LEOP, **global**



Figure 1.2.7: *Large (30 m) S-band antenna at DLR's ground station in Weilheim. (Source: DLR)*

ground station networks are established (Figure 1.2.8). These take over the necessary data traffic in sequence so that gaps between passes over the antenna and resulting gaps in monitoring and control do not get too large.

For ground stations at medium latitudes (e.g., DLR's Weilheim ground station), satellites in polar orbits are visible only four to six times a day when they reach a sufficient height above the horizon (Figure 1.2.9). For ground stations near the terrestrial poles, however, there is visibility with sufficient elevation for every orbit.

In addition to the availability of communication resources the **control loop time** (**latency**) also affects the mission concept. This latency is defined as the overall time for signal transmission and processing in the loop: telemetry generation – telemetry transmission – telemetry processing – command generation – command transmission – command implementation. If the control loop time is longer than the necessary reaction time, autonomous operation of functions becomes mandatory.

1.2.2 Design of System Segments for Space Flight Missions

The starting point for the design of the **complete system** of a space flight mission is the planned objective and the spacecraft's application, as well as the

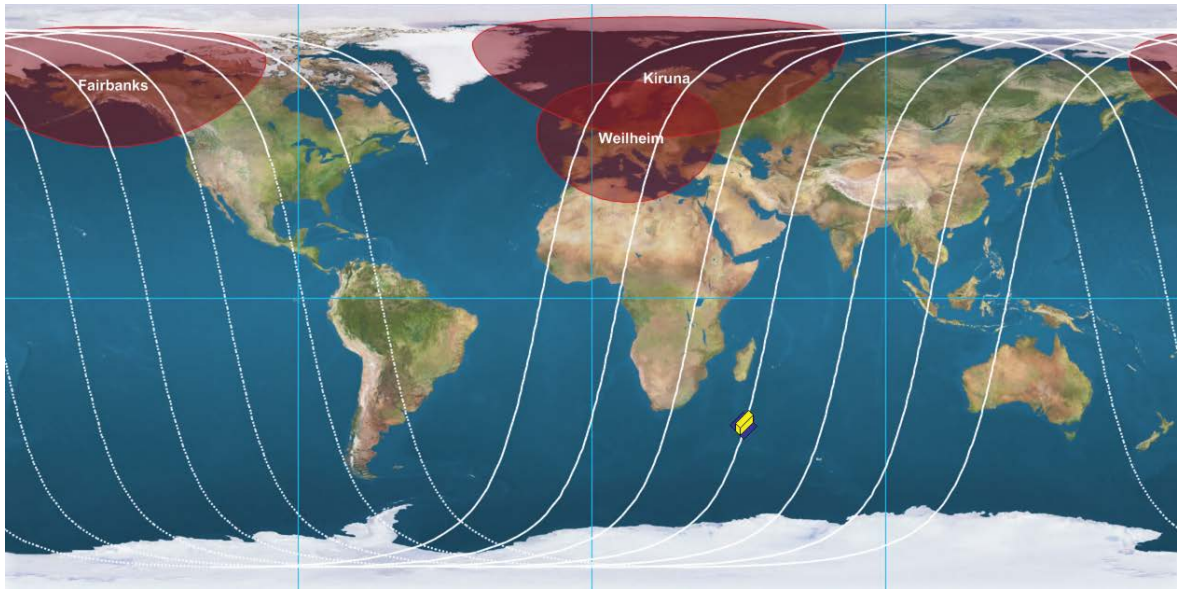


Figure 1.2.8: *Global ground track of the polar orbiting DLR satellite BIRD during the LEOP. Also illustrated are visibility areas of the central DLR ground station in Weilheim and further ground stations belonging to the BIRD-LEOP network at high latitudes (Kiruna, Sweden and Fairbanks, Alaska). With their help the satellite is visible at least once every 90 minutes. Thus telemetry can be downloaded and the satellite controlled by commands 15 times per day (Source: DLR).*

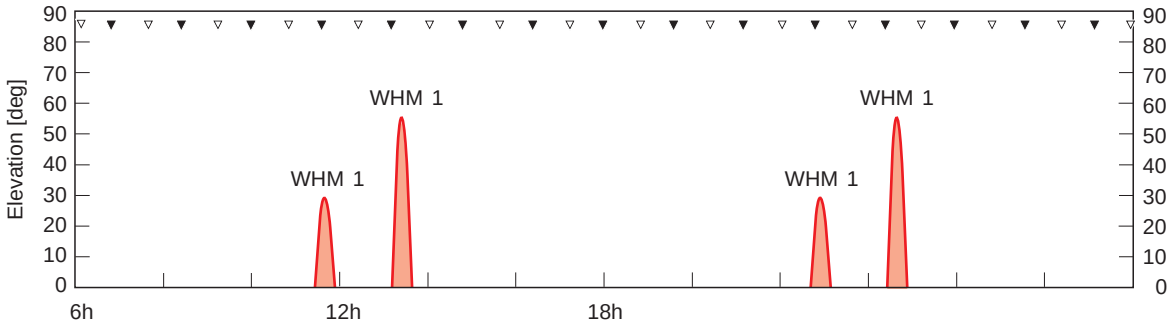


Figure 1.2.9: *Elevation for four BIRD contacts during one day with the DLR Weilheim ground station (corresponding to the two successive passages in Figure 1.2.8). Contacts last less than 10 minutes each. Failure-free data exchange is only possible at an elevation over 5° (Source: DLR).*

resulting requirements (Figure 1.2.10). There is no fixed procedure for the design of a space flight system. Important aspects beyond those described here are covered in Chapter 8. Moreover, the standard publications [1.2.2] to [1.2.6] describe relevant methods of systems engineering.

The mission objective determines the mission. Thus the description of the payload often stands at the beginning of the full system description. During payload design, quantitative properties of the payload, such as optical resolution (Earth observation), transmission capacity (communications) or signal accuracy (navigation) are the center of attention. However, lifetime as well as the interaction of components and operational manageability are also important aspects. Depending on the payload requirements, characteristic parameters (mass, volume, energy demand, communication demand, etc.) are estimated and further requirements for the spacecraft and its orbit (pointing accuracy, ground track and times, measurement geometry) are determined.

The system engineer often works initially in a so-called **top-down design** approach within the hierarchical structure of a system (Figure 1.2.11) on the concepts for the upper levels (beginning with the system elements). In the course of the design process more details become clarified and the lower system levels are defined.

Using mission objectives and payload characteristics it is possible to define an **optimum orbit** applying an adequate mission analysis. This can be a suitable

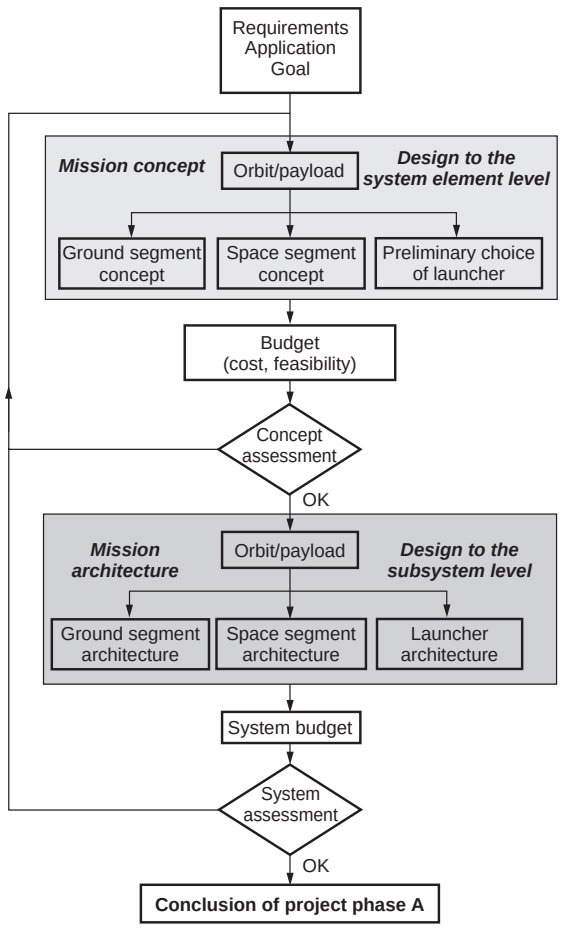


Figure 1.2.10: *Flow chart for defining a space flight mission (project phase A).*

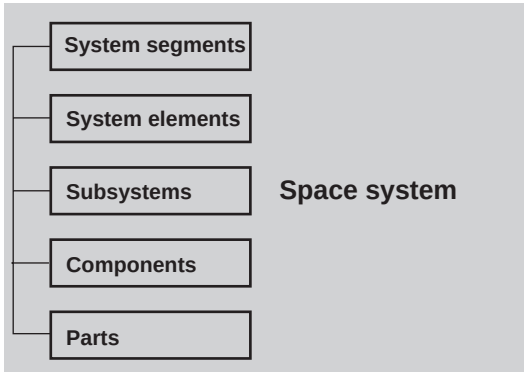


Figure 1.2.11: *Hierarchical structure of a space flight mission.*

Earth orbit as well as an interplanetary orbit or even an orbit around another celestial body. Despite the potential to create optimal solutions in the orbit selection process, further options are helpful, as reasons might emerge during the actual system design for choosing a nonoptimal orbit. Typical examples of such reasons are ground station availability or aspects of the spacecraft's energy supply.

When visions concerning payload and orbit layout have converged, one can start by describing the spacecraft. At first the spacecraft can be described by defining subsystem requirements. A first set of characteristic parameters for the spacecraft and orbit can be defined.

Knowing these characteristic spacecraft data (volume, mass, etc.) and orbit (propulsion demand, location of launch place, etc.) a first selection of an appropriate launcher becomes possible. At the same time spacecraft (communication demand, degree of autonomy, etc.) and orbit (antenna visibility) parameters are used to design the ground segment.

This initial rough knowledge of payload, orbit and spacecraft permits the determination of a suitable launcher and the design of the ground segment, resulting in a preliminary mission concept for the space system.

The first **mission concept** also enables a rough check of superior mission constraints to be made, such as costs and technical feasibility. This presupposes a preliminary understanding of the properties and costs of the potential system segments including spacecraft, launcher and ground segment.

After the first estimation of feasibility and costs the mission concept is revised. If a stable mission concept (or several apparently equal mission concepts) is obtained after a number of iterations, the mission architecture can be built up by developing the subsystem design from the subsystem requirements. This enables detailed cost and feasibility analyses. This basis is often followed by further iterations of the mission architecture until a number of stable alternatives are defined. Within a system analysis the favored alternative is chosen for subsequent realization.

The two essential conditions for the design to converge into an optimized complete system are transparency of the design process and completeness of the system and subsystem aspects. Today, both conditions can be supported by the new methods and tools of **concurrent engineering**, especially for complex missions. Specialists in all relevant domains work on the design process simultaneously, which increases efficiency. Only a clear solution which includes the ground segment and launcher and all their critical aspects can confirm the **feasibility** of a mission. For the launcher, primarily the orbit to be reached, the mass and volume of the spacecraft and its availability, reliability and costs are of outstanding importance.

Following the determination of the mission architecture, which is the subject of a Phase A study, typically a Phase B study or definition study is conducted in order to define the components and individual parts of all subsystems. This is followed by the final cost and feasibility analysis. Design, construction, integration, test and qualification are combined in Phase C/D, which is followed by operations (Phase E) and, finally, secure shutdown and disposal of the space flight system.

In creating a mission concept, useful experience can be formulated as the following rules of thumb:

1. The feasibility of a mission is often dependent on the compatibility of a mission with a **financial budget**. Ground segment costs and the range of operation tasks are dominated by the required infrastructure. The cost of the transfer segment is primarily dominated by spacecraft mass and geometry as well as propulsion demand. Space segment costs are influenced by the complexity

of the spacecraft and the payload. Minimizing mass leads in general to an increase of complexity. Therefore optimization is also necessary here. In all domains experience (available operation processes, experience with launchers, serial production of satellites and payloads) results in distinct cost reductions.

2. The mass of the spacecraft platform is three to ten times the mass of the payload. For large communications satellites this mass ratio is often between three and five. On planetary and in earlier missions the mass ratio favored the platform even more.
3. The mass of the satellite subsystems is dominated by the **power supply** and the **structure**. Together both mass subsystems comprise between one-third and one-half of a spacecraft's dry mass. Propulsion and attitude control are next with regard to the proportion of mass.
4. The **operation effort** varies strongly over the **lifetime of a mission**. The greatest effort is required during the few days of the launch and early operation phase during which the subsystems are adjusted to the extreme environmental conditions of space after ascent and separation from the launcher. After stabilizing the spacecraft's condition there is a commissioning and test phase during which attention is mostly on the configuration of the payload. In this phase the operation effort is still rather high compared to the subsequent routine operations phase. In this phase, sometimes lasting several years, only a fraction of the initial operations crew of the first two phases is required, especially if operation takes place in a so-called **multimission environment**, in which several spacecraft are controlled simultaneously. With increasing mission duration the operation effort also increases again after a few years as components and subsystems subject to deterioration fail. Typically, attitude control and power are affected most seriously. In general, the number of unpredictable failures increases with lifetime, which can lead to the destabilization of the entire spacecraft. The operation effort increases dramatically again and often stays at a relative high level until the end of the mission. This operations profile from launch to end of life is termed an asymmetric "bathtub profile."

1.2.3 Space Flight Mission Classification

The decisive factor for the properties of a spacecraft, the corresponding ground segment and the mission is its application and the corresponding requirements. In addition to standard systems for the spacecraft, such as standard platforms with different payloads, there are also standardized components for ground

Example

TerraSAR-X:	*German radar Earth observation satellite (Figure 1.2.12)*
Launch:	*June 2007*
Mass:	*1200 kg*
Altitude:	*520 km*
Inclination:	*97°*
Launch site:	*Baikonur*
Payload:	*High-resolution X-band radar*



Figure: 1.2.12: *TerraSAR-X satellite during acceptance tests before transport to the launch site. The payload (X-band radar) is integrated into the main body of the satellite. Communication occurs via a boom antenna, which can be seen on the right in its folded state (Source: DLR).*

segment equipment and the launcher. There are nine **characteristic areas of application** for space flight missions, as follows.

1.2.3.1 Earth Observation

Earth observation missions are conducted with small as well as medium and large satellites. Earth observation comprises data acquisition with cameras and sensors which work at different wavelengths over infrared, visible and ultraviolet ranges. Both passive optical observation techniques and active sampling (radar) are used. Moreover, a variety of different measuring methods are applied to determine the electric, magnetic, optic or gravimetric properties of the Earth and its atmosphere. Orbits for Earth observation are often rather low (< 1000 km) and have a high inclination in order to be able to observe the Earth with high resolution and from pole to pole. Sometimes a target is to be imaged at a certain time of the day (to achieve equivalent shadowing conditions). That requires a rotation of the orbital plane of approx. 1° around the Earth's axis every day (**Sun-synchronous orbit**). This rotation can be generated by an orbit disturbance caused by the Earth's oblateness. It requires that the orbit is not exactly above the poles, but slightly inclined (as in the TerraSAR-X mission). In case the target is to be reanalyzed under the same observation angle, the satellite has to be brought onto the original track after a few orbits (**repeat ground track**).

1.2.3.2 Weather Observation

A special case of Earth observation is weather observation. Since the beginning of space flight missions weather observation has been a continuously expanding area of application with numerous satellites in low as well as geostationary orbits. Central aspects are the local, regional and global analysis of weather conditions, the generation of data as input for weather forecast models, and the characterization of the atmosphere regarding properties and changes in regional and global climate. Imaging instruments are often used on the satellites. Commercial user scenarios in weather observation are also gaining in importance.

Example

MetOp-A:	*European weather satellite (Figure 1.2.13)*
Launch:	*October 2006*
Mass:	*4093 kg*
Altitude:	*820 km*
Inclination:	*99°*
Launch site:	*Baikonur*
Payload:	*13 instruments for weather observation*



Figure 1.2.13: *The European MetOp-A satellite is in a Sun-synchronous LEO. The satellite operates together with the American NOAA weather observation satellites. This suits the purpose of optimizing the coverage of relevant observation areas (Source: ESA).*

1.2.3.3 Technology Testing

Technology testing missions are used for testing and validating technical components and procedures under space conditions. On the one hand this can be done for satellite components and payloads for operational applications as well as new materials or robotic components. On the other hand new communication and navigation procedures can be tested. Technology tests are intensively conducted also in the context of human missions. Since the 1970s space laboratory modules

Example

BIRD:	*German technology testing satellite (Figure 1.2.14)*
Launch:	*October 2001*
Mass:	*92 kg*
Altitude:	*580 km*
Inclination:	*97°*
Launch site:	*Sriharikota*
Payload:	*Infrared detectors, WAOSS camera system, GPS navigation system, etc.*

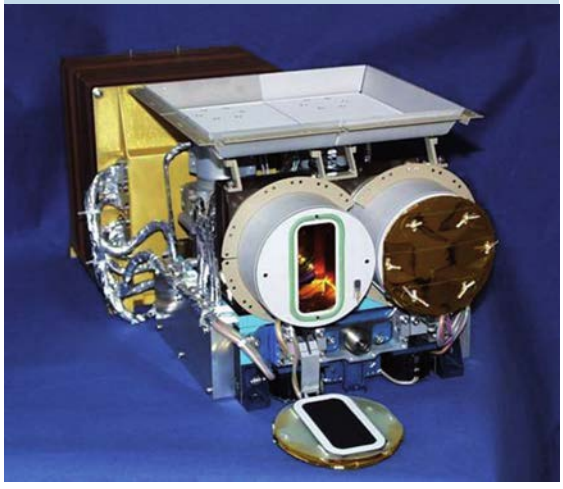


Figure 1.2.14: *BIRD payload: a two-channel infrared camera as part of an extensive technology test package on the DLR satellite (Source: DLR).*

have been used. Beside platforms such as the ISS, small spacecraft and sometimes microsatellites with a mass of less than 100 kg in low orbits are also employed.

1.2.3.4 Fundamental Research

Fundamental research missions typically suit the purpose of studying **astronomical objects** or **physical phenomena** in the context of cosmology or analyses in relativistic physics. The range of instruments used for that purpose covers the complete electromagnetic spectrum as well as precise experimental arrangements. Space telescopes are also used and can be extremely large and complex. Interlinked systems comprising several satellites gain more and more importance. Orbits for such missions are very diverse, and in some cases have extreme altitudes (100 000 km), for example to avoid magnetospheric disturbances.

Example

ROSAT:	*German X-ray telescope satellite (Figure 1.2.15)*
Launch:	*June 1990*
Mass:	*2426 kg*
Altitude:	*570 km*
Inclination:	*58°*
Launch site:	*Cape Canaveral*
Payload:	*Four-times nested Wolter telescope*

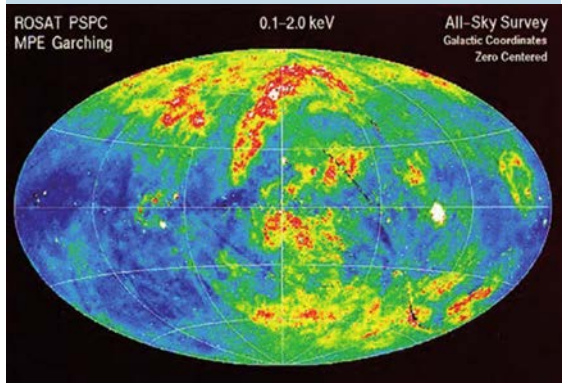


Figure 1.2.15: *Complete image of the sky in X-ray light (0.1–2.0 keV) from ROSAT; observation of the first half year of almost 10 operating years (Source: MPG).*

1.2.3.5 Communication

Communication is by far the greatest area of application for the **commercial operation** of satellites. The required electrical power and the typical propulsion requirements for the positioning of these satellites in geostationary orbit lead to great dimensions and several tonnes of total mass. These satellites are launched into a so-called geostationary transfer orbit (perigee approx. 500 km, apogee approx. 36 000 km) by the most powerful launchers available. The adaptation of this orbit into the final geostationary orbit occurs by so-called **apogee boost maneuvers** with the satellite's main engine in the first few days of the mission. Due to the large number of satellites in geostationary orbit with its subdivision into discrete control boxes, the demand on flight dynamics is continuously high. The density caused by having hundreds of communications satellites located in this orbital region makes it necessary to put the satellite into a so-called graveyard orbit some hundreds of kilometers beyond at the end of its service life (nominally about 15 years). There are also communications satellites, for example as constellations (Iridium), in lower orbits.

1

Example

EUTELSAT W5:	*Communication satellite of the European operator EUTELSAT (Figure 1.2.16)*
Launch:	*November 2002*
Mass:	*3170 kg*
Altitude:	*35 000 km*
Inclination:	*0°*
Launch site:	*Cape Canaveral*
Payload:	*24 Ku-band transponder*

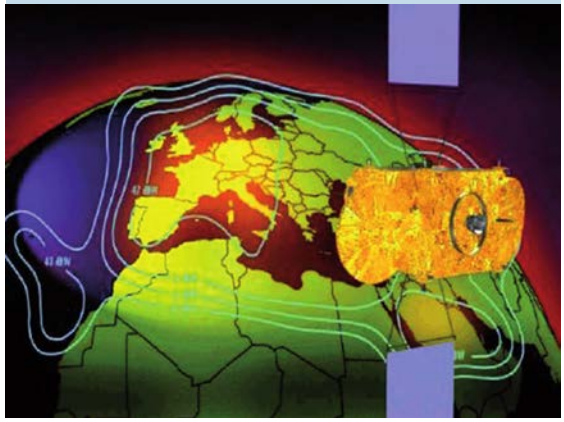


Figure 1.2.16: *Schematic illustration of the so-called antenna mapping output of a geostationary communications satellite. The irregular form of the lines of equal irradiation intensity is exactly controlled by an appropriate antenna design. At the end of the positioning procedure the satellite's radiation power profile is measured on the ground by changing the orientation of the satellite. Subsequently, the satellite is transferred into routine operations (Source: DLR).*

Example

Galileo:	*European navigation satellite system (Figure 1.2.17)*
Launch:	*2010–2013*
Mass:	*680 kg*
Altitude:	*23 600 km*
Inclination:	*56°*
Launch site:	*Kourou*
Payload:	*Navigation signal transmitter and high-precision clock*

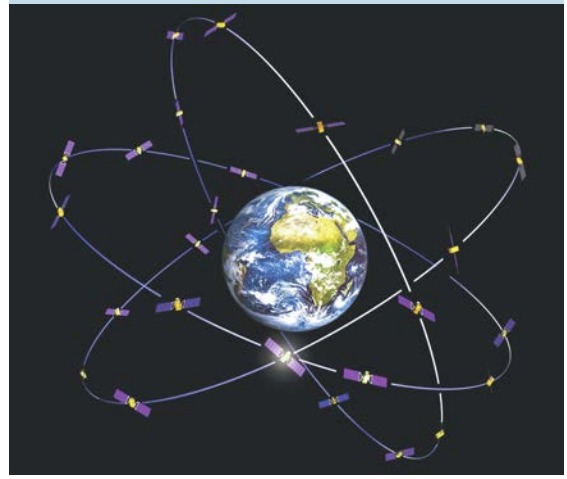


Figure 1.2.17: *Schematic illustration of the Galileo constellation with 27 operating satellites and three spare satellites on standby. Monitoring and control of the satellites are carried out at control centers in Germany (Oberpfaffenhofen) and Italy (Fucino) (Source: ESA).*

1.2.3.6 Navigation

With increasing utilization of the US **global positioning system** (GPS) since the 1970s, the significance of this application has grown rapidly. Navigation satellites provide a permanent signal by which, using at least four satellites at the same time, any position on Earth can be determined. Orbits are typically medium high; in the case of GPS, 20 183 km. In order to provide a global navigation service, there have to be more than 20 identical satellites in a constellation. In addition to other systems the European **Galileo navigation system** will gain outstanding significance as a commercial system.

1.2.3.7 Military Missions

Military missions comprise mission types with applications in the areas of communication, navigation, Earth observation, technology testing and weather observation. However, the employed spacecraft differ in data protection (encoding), general secrecy and in many cases hardening against electromagnetic interference. In addition, there may be a small number of satellites with interfering or destructive functions. This huge variety of military spacecraft results in an equally high variety of spacecraft and orbits. In the area of high-resolution Earth observation, very low orbits dominate and result in relatively short lifetimes for the satellites. In Germany satellite missions are being conducted in this area of application for the first time with the SAR-Lupe reconnaissance constellation.

1.2.3.8 Planetary Exploration

To this day planetary exploration has remained an activity of unmanned space flight except for the Apollo

Example

SAR-Lupe:	*German radar intelligence satellite constellation (Figure 1.2.18)*
Launch:	*December 2006*
Mass:	*770 kg*
Altitude:	*490 km*
Inclination:	*98°*
Launch site:	*Plesetsk*
Payload:	*High-resolution X-band radar*



Figure 1.2.18: *The satellites of the German SAR-Lupe constellation used for radar intelligence (Source: OHB System).*

Example

Rosetta:	*European comet probe (Figure 1.2.19)*
Launch:	*March 2004*
Mass:	*3100 kg*
Altitude:	*In planetary orbit with several Earth, Mars and asteroid fly-bys*
Launch site:	*Kourou*
Payload:	*Lander with drill, cameras, spectrometer, etc.*

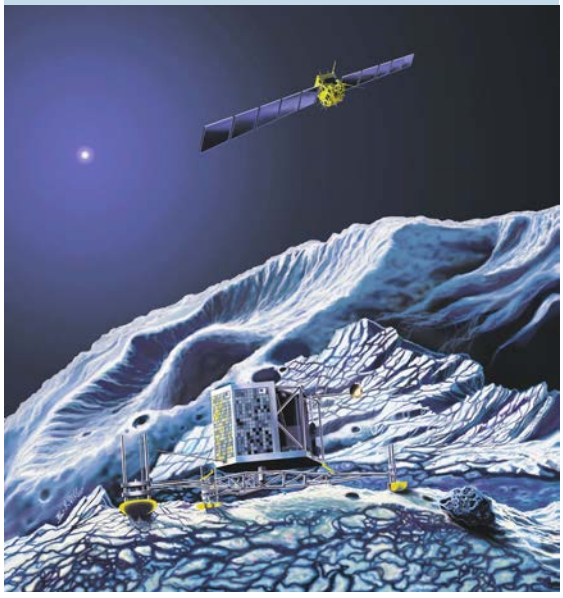


Figure 1.2.19: *Conceptual illustration of a successful landing maneuver on the comet 67P/Churyumov–Gerasimenko in 2014 (Source: ESA).*

missions to the Moon. Over the decades dozens of **space probes** have flown to all planets (except Pluto) as well as to many asteroids and comets. In many cases these were not only fly-bys but even orbiting and landing missions. For instance, robotic vehicles have been deployed on Mars, or cometary material has been returned to Earth. Special challenges of **interplanetary space flight** are significant signal latency (up to hours in the outer Solar System), long flight time, and navigation. Moreover, a sufficient power supply is one of the greatest problems in the outer Solar System.

1.2.3.9 Human Space Flight

The unique requirements of **life support** mean a considerably higher effort in the area of human space flight. Basically one has to distinguish between **transport systems** like the Space Shuttle or Soyuz spacecraft and **long-term orbiting systems** like the ISS. One strives to reduce the flight and maintenance effort in favor of science and experiment time. Human space flight in Earth orbit is typically realized with medium inclinations at relatively low altitudes. The spacecraft typically have much more mass than the satellites.

Bibliography

[1.2.1] Union of Concerned Scientist Satellite Database. http://www.ucsusa.org/global_security/space_weapons/satellite_database.html, 2007.

[1.2.2] Brown, Ch. *Elements of Spacecraft Design*, AIAA Education Series. Reston, VA: AIAA, 2002.

1

[1.2.3] Fortescue, P.W., Stark, J.P.W., Swinerd, G. *Spacecraft Systems Engineering*. Chichester: John Wiley & Sons, Ltd, 2003.
[1.2.4] Griffin, M.D., French, J.R. *Space Vehicle Design*, AIAA Education Series. Reston, VA: AIAA, 2004.
[1.2.5] Pisacane, V.L. *Fundamentals of Space Systems*, Johns Hopkins University/Applied Physics Laboratory Series. New York: Oxford University Press, 2005.
[1.2.6] Wertz, J.R., Larson, W.J. *Space Mission Analysis and Design*, Space Technology Library. Dordrecht: Springer Netherland, 1999.

Example

Columbus:	*European space laboratory on the ISS (Figure 1.2.20)*
Launch:	*February 2008*
Mass:	*10 275 kg*
Altitude:	*350 km*
Inclination:	*51°*
Launch site:	*Cape Canaveral*
Payload:	*Internal and external experiment modules, racks and crew*

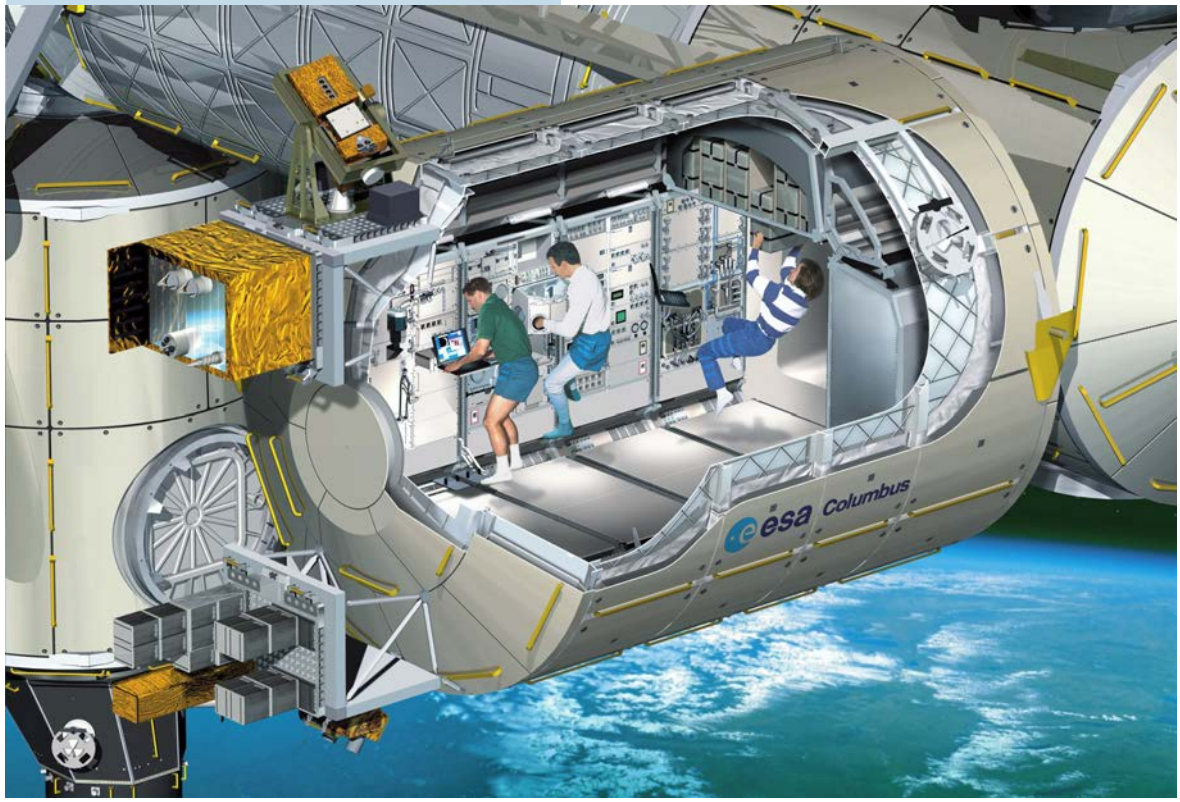


Figure 1.2.20: *The European Columbus laboratory module on the ISS operated from Oberpfaffenhofen (Source: ESA).*



2 Fundamentals

Klaus Wittmann

The design of a space system requires an in-depth knowledge of the internal interfaces between the transfer segment and the space segment (volume, mass, static and dynamic loads) and between the space segment and the ground segment (operating concept, communications). These interfaces will be thoroughly covered in Chapters 3 through 7.

This section describes the **interfaces** of a space system with its environment and the laws describing the motion of a spacecraft.

A spacecraft is exposed to the Earth's atmosphere until it reaches the environment of space after launch. The conditions on the ground are familiar to us. Thus, we will not give here any details of atmospheric conditions on the ground, although they have some influence, for example as cleanliness conditions during integration of a spacecraft or as weather conditions at launch.

The **space environment**, especially the influence of the Sun and the Earth, is characterized by:

- High-energy electromagnetic radiation
- Particle radiation
- Vacuum
- Low temperature of the space background
- Reduced gravity
- Residual atmosphere (in low orbits).

An overview of the characteristics of the space environment and their effects on a spacecraft and its payload is given in Section 2.1.

The **motion of a spacecraft** is governed by the laws of orbit dynamics. Section 2.2 describes the methods of orbit determination and prediction. Typical orbit maneuvers of the spacecraft as well as orbit perturbations are discussed. Characteristic orbits for typical mission profiles are presented. The relatively complex mathematical elaboration of the governing laws cannot be provided within the scope of this book. However, the mathematical formulas which are used in practical applications are given and the methods to apply them are described.

When a spacecraft returns to Earth the influence of the atmosphere gains in importance. Section 2.3 covers the field of **aerothermodynamics**, which is relevant for this phase of a space mission, especially as to the design of a return vehicle.

Section 2.4 returns to the space environment in order to present one specific aspect, namely **micrometeorites** (natural causes) and **space debris** (objects and particles introduced by space flight activities), which represent a hazard for spacecraft. Thus, Section 2.4 characterizes the micrometeorite and space debris environment and gives an overview of the models employed to quantify the risks. Strategies for avoiding space debris and protective measures are also described.

2.1 The Space Environment

Christian Henjes, Holger Kügler, Wilfried Ley, Manfred Magg and Steffen Scharfenberg

2.1.1 Spacecraft and the Space Environment

A spacecraft is affected by physical conditions in space which go far beyond the well-known environmental requirements on Earth. Typical features of the space

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9

2

environment are **high vacuum**, short-wave **solar radiation** (electromagnetic waves), **ultraviolet X-rays and gamma radiation** from the galactic background, **high-energy particles** (electrons, protons, neutrons and alpha particles), the **cold background** of space, **microgravity, aerodynamic drag** of the atmosphere at low Earth orbits, and the influence of **atomic oxygen.** These conditions have to be considered during the design and realization of a spacecraft.

The atmosphere surrounding a spacecraft is mainly characterized by temperature, density, velocity distribution and composition and can be divided into:

- Natural (physical) atmosphere.
- Atmosphere induced by the spacecraft.
- Atmosphere influenced by other space flight missions.

Especially the natural conditions depend on the orbit of the mission. The main distinctions are:

- Low Earth orbit (LEO)
- Medium Earth orbit (MEO)
- Geostationary orbit (GEO)
- Polar orbits (e.g., Sun-synchronous, Molniya orbit)
- Highly eccentric orbits (HEO, GTO)
- Orbits around the Lagrange points
- Interplanetary space trajectories
- Planet orbits and conditions for landing, ascent and ground operations.

The **natural atmosphere** (Earth's atmosphere) moves relative to the spacecraft with an average velocity v_R (e.g., $v_R = 7800$ m/s) and an average thermal velocity $\overline{c}$ (e.g., $\overline{c} = 1100$ m/s at an orbital altitude of 300 km). The **dynamic pressure** which builds up at the nose

$$p_{dyn} = \left(\frac{v_R}{\overline{c}}\right)^2 \cdot p_{stat} = 50.3 \cdot p_{stat} \qquad (2.1.1)$$

therefore exceeds the static pressure by a factor of 50, whereas a wind shadow region arises at the tail since there the dynamic pressure p_{dyn} is much lower than the static pressure p_{stat}.

The radiation conditions for a spacecraft – even during an interplanetary mission – are essentially defined by the influences of the Sun (Figure 2.1.1) which can be divided into:

- Solar wind as ionized gas composed of protons and electrons, which are shielded by the Earth's magnetic field – except at the polar caps.
- Magnetic fields from solar flares.

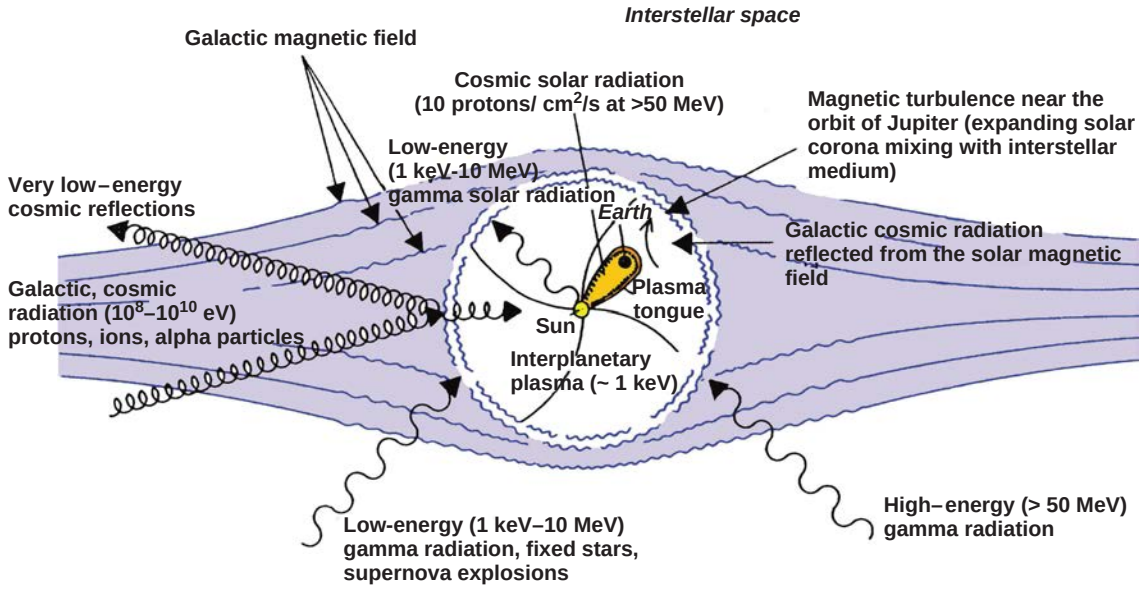
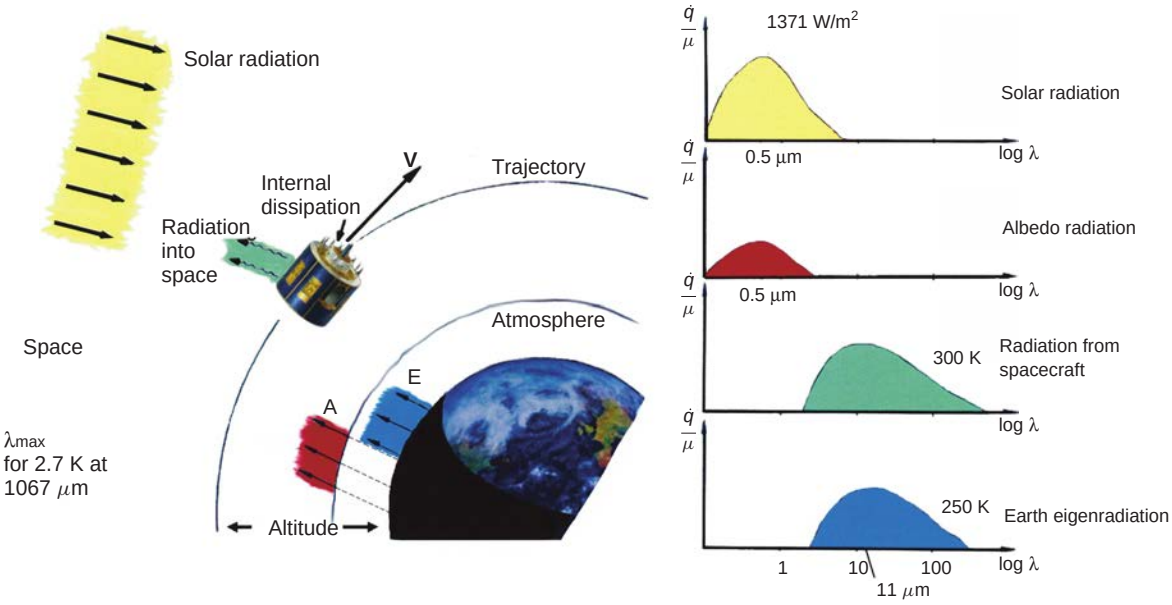


Figure 2.1.1: *Radiation fields [2.1.15].*



E = Earth eigenradiation, dependent on the distance from the Earth
A = albedo, the solar radiation reflected from the Earth, dependent on the distance from the Earth and the angles of inclination and declination

Figure 2.1.2: *Environmental conditions for a spacecraft in LEO [2.1.15].*

- Radioactive radiation.
- Solar radiation (solar constant $SC = 1371\ W/m^2$) as electromagnetic waves which essentially influence the thermal control.

The **induced atmosphere** constitutes a dragging gas cloud produced by the spacecraft itself and expands into space, which acts as a pump with unlimited vacuum capability. This gas cloud is caused by:

- Outgassing of spacecraft parts
- Leakage of pressure tanks
- Exhaust gases from steering and propulsion systems for orbit and attitude control and for maneuvers.

The induced atmosphere mainly consists of water vapor and high-molecular substances from synthetics with condensation probabilities close to the value of 1. In LEOs at 200–300 km they can collide with the gas particles of the incident flow in the direct environment of the spacecraft, partially scatter back on the spacecraft, and thereby contribute to the self-contamination of optically sensitive components (solar cells, sensors, etc.). The environmental conditions for a spacecraft in LEO are shown in Figure 2.1.2.

2.1.2 Influence of the Sun and the Space Background

2.1.2.1 Solar Physics

The physics of interplanetary space is significantly dominated by the influence of the Sun. Therefore its physical characteristics are summarized here first (see Table 2.1.1).

Nuclear fusion processes occur in the center of the Sun at a temperature of $15 \cdot 10^6$ K. In the most important process four protons fuse into one ^{4}He nucleus in several steps. This process frees energy of 26.3 MeV and two neutrinos per ^{4}He core. The positrons that are additionally generated during these reactions are destroyed by the electrons of the plasma and thereby contribute to the release of energy.

Table 2.1.1: *Characteristics of the Sun.*

Radius	$6.96 \cdot 10^5$ km $= R_S$
Mass	$1.99 \cdot 10^{30}$ kg
Particles within plasma	91% H, 8% He, 1% other
Mass density on average	$1.41 \cdot 10^3$ kg/m^3
Mass density at center	$1.5 \cdot 10^5$ kg/m^3
Luminosity	$3.86 \cdot 10^{26}$ W
Effective radiation temperature	5780 K
Solar constant	1.371 kW/m^2 ± 0.3%
Distance to Earth on average	$1.496 \cdot 10^8$ km = 1 AU (Astronomical Unit)
Distance to Earth at perihelion	$1.471 \cdot 10^8$ km
Distance to Earth at aphelion	$1.521 \cdot 10^8$ km
Sidereal rotation period at the equator	24.8 days

The electrically neutral neutrinos leave the Sun at the speed of light and have no significance for astronautics due to their weak interaction with matter.

The region of the core of the Sun extends from its center out to a radius of approximately $2.5 \cdot 10^5$ km, followed by a sphere or shell extending up to $5 \cdot 10^5$ km. The energy generated within the core region is transmitted by radiation transport through the solar plasma of the shell in the direction of the limb of the Sun. During its transport the energy of fusion in the form of hard gamma radiation is continuously absorbed and reemitted and thereby converted into heat (thermalized). On the last part of the journey, from $2 \cdot 10^5$ km up to the limb of the Sun, the transport of heat is by convection.

The atmosphere surrounding the limb of the Sun (Table 2.1.2) is divided into three spheres:

- Photosphere
- Chromosphere
- Corona.

The **corona** has no sharply defined upper border. Above approximately six solar radii it merges into the solar wind of interplanetary space (heliosphere).

2.1.2.2 Solar Radiation

Most of the **radiation energy** is emitted by the photosphere, which is only 200 km thick and, from the Earth's point of view, optically perceived as the solar disc.

Table 2.1.2: *Characteristics of the solar atmosphere.*

	Temperature [K]	Height [km]	Emitted spectrum
Photosphere	3900–7160	0–200	UV, visible light, NIR
Chromosphere	42 000–10 000	200–2500	UV, Hα
Corona	$1 \cdot 10^6$–$2 \cdot 10^6$	2500–6R_S	EUV, X-ray, radio

The spectrum of the Sun, ranging from near **ultraviolet** up to the **infrared** area, can roughly be described by **Planck's radiation law** for a **cavity emitter**. At a distance of 1 AU from the limb of the Sun the spectral density of the energy flow is given by

$$S_S(\lambda)\big|_{1\,\mathrm{AU}} = \left(\frac{R_S}{1\ \mathrm{AU}}\right)^2 \cdot 2\pi \cdot h \cdot c_0^2 \cdot \frac{1}{\lambda^5} \cdot \frac{1}{e^{\left(\frac{h \cdot c_0}{\lambda \cdot k \cdot T}\right)} - 1} \tag{2.1.2}$$

In this formula:
R_S = current Earth–Sun distance,
λ = the wavelength of the radiation,
$h = 6.6261 \cdot 10^{-34}$ J s, Planck's constant,
$k = 1.3807 \cdot 10^{-23}$ J/K, Boltzmann's constant,
T = temperature of radiation,
$c_0 = 2.9979 \cdot 10^8$ m/s, the speed of light in vacuum.

The energy flux density integrated over all wavelengths at 1 AU distance to the Sun is called the **solar constant** (*SC*). From measurements not influenced by the Earth's atmosphere the following numerical value can be determined:

$$1\ SC = \int_0^\infty \mathrm{d}\lambda \cdot S_S(\lambda)\,\big|_{1\ \mathrm{AU}} = 1371\ \frac{\mathrm{W}}{\mathrm{m}^2}\ (\pm 0.3\%) \tag{2.1.3}$$

At a distance of 1 AU from the Sun the radiation is collimated. The radiances differ from the line between the solar center and the point of observation by ±0.27° maximum. For **collimated radiation** the spectral density of the energy flux is defined by the radiation power per

interval of wavelength falling on a unit of area perpendicularly oriented to the direction of the radiation.

The **energy flux density** (Equation 2.1.2) received from the Sun at any location in space has to be scaled by the square of the distance between this location and the Sun. The seasonal distance variation between perihelion and aphelion implies a corresponding variation in the solar flux density on Earth. The flux density on Earth on day n of a year, counting from perihelion crossing (January 3), is approximately equal to

$$S_S \approx S_S\big|_{1\,AU} \cdot \left[1 + 0.033 \cdot \cos\left(2\pi \cdot \frac{n}{365}\right)\right] \quad (2.1.4)$$

There are two ways to define the solar **radiation temperature**. Using the "Wien shift law,"

$$\lambda_{max} \cdot T = 2897\ \mu m \cdot K \quad (2.1.5)$$

the temperature of radiation is defined by the wavelength λ_{max} at which the spectral energy flux density is a maximum. For sunlight this maximum is at 0.45 μm (blue). This results in a temperature of $T = 6400$ K.

However, if the Stephan–Boltzmann law is used to determine the energy flux of a cavity emitter summed up over all wavelengths

$$S_S^{total} = \sigma \cdot T^4 \quad (2.1.6)$$

with

$$\sigma = 5.67 \cdot 10^{-8}\ W/(m^2\ K^4)$$

and if the density of energy flow is equated with the illuminating power of the Sun of $6.34 \cdot 10^7$ W/m^2 relative to the solar surface, this results in the so-called effective radiation temperature $T_{eff} = 5780$ K(see Figure 2.1.3).

The fact that these two definitions of solar radiation temperature lead to different results is because the solar spectrum is not exactly of the type of a cavity emitter.

The effective radiation temperature is better suited for assessing the thermal effects of solar radiation on spacecraft.

The solar constant (at a distance of 1 AU from the Sun) varies temporally by less than 0.3%.

The contribution to the solar radiation of the extreme ultraviolet and the soft X-ray field ($\lambda < 0.2$ μm), which is insignificant if taken absolutely and comes

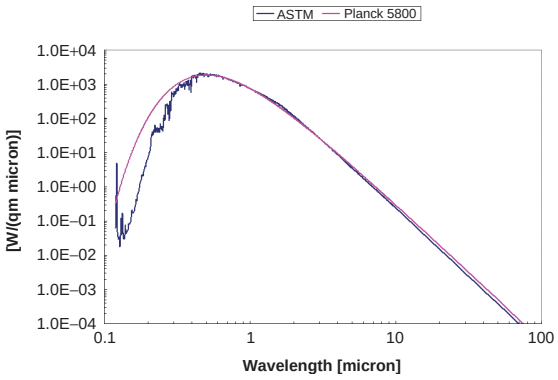


Figure 2.1.3: *Spectral density of energy flow of solar radiation for a distance of 1 AU from the Sun. Comparison of ASTM standard spectrum with black temperature beam at 5800 K.*

primarily from the chromosphere and corona, does vary significantly with the 11-year solar cycle and the rotation period of the Sun. This portion of the radiation is absorbed by the Earth's thermosphere where it has a strong impact on the temperature and therefore on the density profile in the upper atmosphere. This effect has to be taken into account during the slowing down of spacecraft in LEO. In connection with eruptions within the chromosphere, called "flares," there is strong emission of radiation in the radiofrequency band. Sometimes even X-rays ($\lambda < 10^{-2}$ μm) are emitted, reaching the Earth's orbit with an energy flux density of 10^{-6} to 10^{-4} W/m^2.

An immediate consequence of solar irradiation of the Earth is the so-called **albedo**. It is created by the diffuse reflection of solar radiation by the part of the Earth that the Sun is shining on. In a thermal analysis it is possible to assume that the spectral distribution of the reflected radiation from the Earth is the same as from the direct radiation of the Sun. The reflection rate varies between 0.05 and 0.6 depending on what part of the Earth can be seen from the spacecraft at a given point of time (clouds, continents, oceans, etc.). For most spacecraft a temporal average of 0.3 can be assumed.

The **eigenradiation** of the Earth is at least partly induced by solar radiation. It is defined as the thermal radiation of the Earth required to achieve thermal equilibrium between the absorbed radiation of the Sun and the heat generated within the Earth itself, coming, for example, from radioactive sources within

2

its core. Also the intensity of this radiation depends on the part of the Earth's surface seen from space. It varies between 150 and 350 W/m^2. The average is assumed to be 230 W/m^2. This equals the radiation of a black radiator with a temperature of 250 K.

Both albedo and eigenradiation are not collimated radiation in those areas where they have to be taken into account for thermal reasons.

2.1.2.3 Solar Wind

Apart from electromagnetic radiation, the Sun also emits material called the solar wind (Table 2.1.3). During this process the Sun loses approximately 1 million tonnes of hydrogen per second. The **solar wind** is a neutral plasma current consisting mainly of protons and electrons. Close to the Earth and more generally speaking in the plane of the ecliptic it can be divided into a slow and a fast mode. The space close to the Earth is alternately swept across by slow and fast solar winds depending on the rotation period of the Sun.

The plasma of the solar wind is cold in the sense that the energy of its flux is much higher than its thermal energy. The solar wind hits the magnetopause of the Earth on the solar side of the magnetosphere, traveling along the magnetic field lines of the Sun into interplanetary space (Figure 2.1.4).

Table 2.1.3: *Properties of the solar wind in the plane of the ecliptic at a solar distance of 1 AU.*

Chemical composition	96% protons, 4% He^{++} (fluctuating), electrons
Density	6 per cm^3 (protons = electrons)
Proton flux	$3 \cdot 10^{12}$ m^{-2}s^{-1}
Temperature	3500–500 000 K
Free path length	10^8 km
Velocity	200–400 km/s (slow) 600–2000 km/s (fast)

By surrounding the magnetosphere of the Earth, the solar wind transfers some of its kinetic energy to the tail of the magnetosphere, which is located on the night side of the Earth and reaches far into interplanetary space. This collected magnetic field energy is discharged from time to time by so-called **magnetic substorms**. During a substorm hot plasma is generated in the tail of the magnetosphere. This plasma is partly ejected into the region of the geostationary orbit.

In addition to the permanent solar wind flowing from the corona of the Sun, a **coronal mass ejection** sometimes occurs. During this process up to 10^{10} tonnes of plasma are ejected into interplanetary space within a few hours. If this plasma cloud crosses the

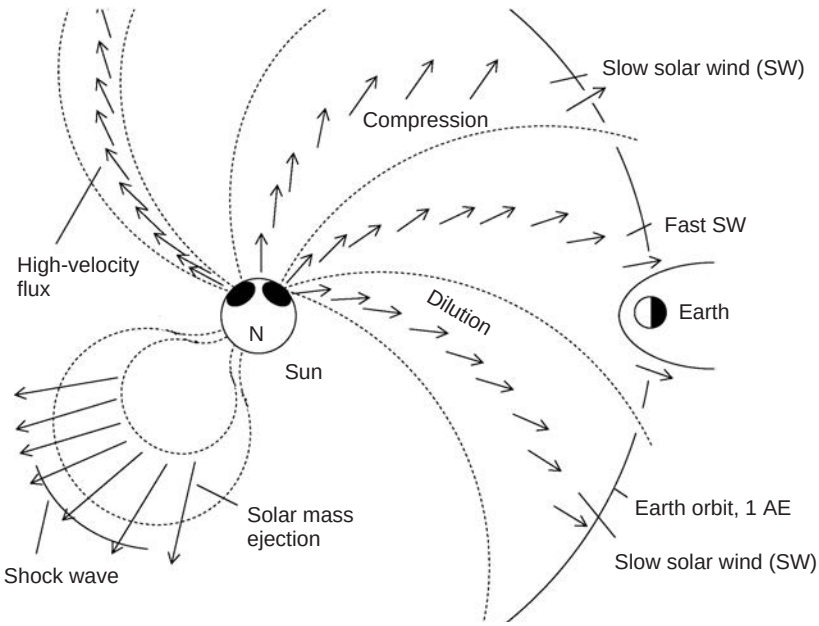


Figure 2.1.4: *Solar wind and interplanetary magnetic field within the ecliptic [2.1.1].*

Table 2.1.4: *Dimension of proton flux within cosmic radiation.*

Energy [MeV]	1	10	100	1000	10 000	100 000
Flux [$m^{-2} \cdot sr^{-1} \cdot s^{-1} \cdot MeV^{-1}$]	10^{-2}	0.1	1	2	10^{-2}	10^{-5}

Earth's orbit, extremely strong magnetic storms are the result.

2.1.2.4 High-Energy Particles

High- and very high-energy particles (mainly protons) reach our Solar System from all directions (i.e., independent of the position of the Sun). Their energy level typically ranges from 100 MeV up to several GeV (Table 2.1.4). Single events at even 10^{20} eV have been observed.

In conjunction with a coronal mass ejection protons with energies of 1 to more than 100 MeV are ejected from the Sun's atmosphere. These eruptions are located on only small areas of the solar atmosphere and last no more than a few hours. Therefore, the ejected protons reach the environment of the Earth only if the magnetic field lines of the Sun that are coming from the area of eruption sweep across space close to the Earth.

2.1.3 Influence of the Earth

2.1.3.1 The Earth's Atmosphere

Earth's atmosphere can be subdivided into several realms based on temperature: the troposphere, stratosphere, mesosphere and thermosphere (see Figure 2.1.5). The upper edge of each layer is always called the same as the layer itself, "pause" instead of "sphere".

Meteorological weather with its highs and lows, including its weather fronts, plays a major role in the troposphere. The dynamics and thermodynamics of the troposphere are essentially influenced by the nearby surface and rotation of the Earth. In addition, water in all its states of aggregation is very significant in the troposphere. The **stratosphere** above, however, is dry. Its diffusion in the vertical direction is much smaller than in the case of the troposphere. Also, it includes the ozone layer. The absorption of radiation within the ozone layer leads to a temperature increase up to a height of 50 km.

In the **mesosphere**, from 50 to 80 km altitude, the temperature decreases. This temperature decrease is accompanied by a smaller pressure decrease with increasing height than would be expected with constant or even increasing temperature. The **thermosphere** which

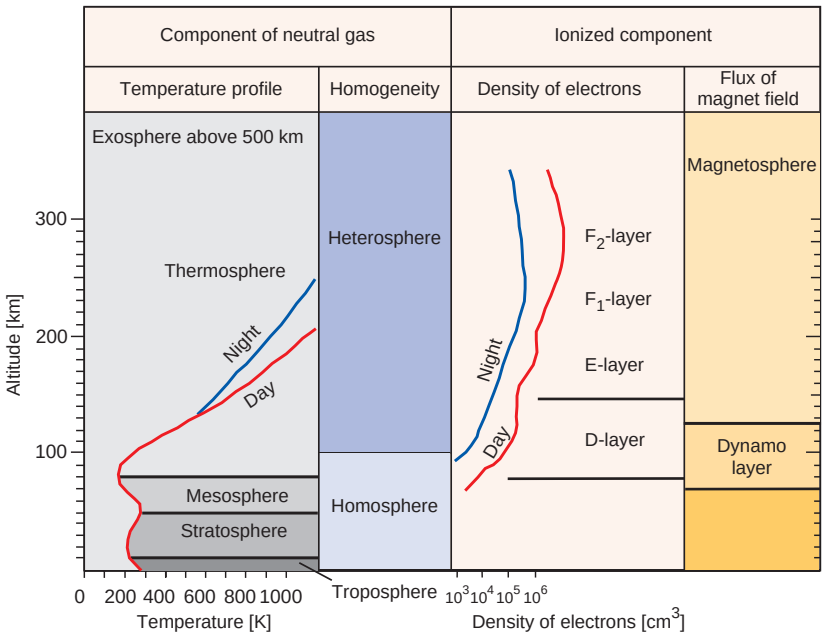


Figure 2.1.5: *Classification of the atmosphere (schematic).*

2

follows is a region where temperatures again increase with altitude. Above 130 km a clear difference in temperature and pressure is observed between day and night.

Using a different system of characterization the Earth's atmosphere can be divided into the **homosphere** and the **heterosphere**. The neutral gases are well mixed up to a height of 100 km. As a direct result of turbulence, the diffusion coefficients of the different components of the air are constant. Accordingly, this layer is often named the **turbosphere** and its upper limit the turbopause.

The layer above is the heterosphere. The neutral gases in this layer are separated by the Earth's gravity in such a way that the percentage of lightweight gases increases with altitude, while the percentage of heavyweight gases decreases. Eventually, the density of the gases is so low that single particles (molecules) can move in Keplerian orbits without colliding with each other. Particles which have a higher than orbital velocity can escape into interplanetary space. This region, beginning at a height between 500 and 600 km, is called the **exosphere**. Only neutral gas particles can leave the gravity field of the Earth, however. Ionized particle behave completely differently.

There are two classification systems for ionized particles in the atmosphere. One is based on the concentration of electrons as a function of altitude; the other is based on the influence of the Earth's magnetic field on the ionized components of the air.

Charged particles that move across the magnetic field are influenced by a force perpendicular to the magnetic field and to their velocity. Due to their smaller mass electrons are more influenced by this force than are ions traveling at the same velocity. The proportion of ionized particles increases with height. Up to an altitude of 70 km the share of ionized particles is so small that the neutral gas particles take the ionized gas particles with them. Above there is a region where the ions completely follow the movement of the neutral gas particles, while the electrons are hindered in their movement across the magnetic field. As a result, electric fields are generated (**dynamo layer**).

The density of the neutral gas decreases above a height of 130 km, which makes collisions between ions and neutral gas particles very rare. The motion of all ionized particles is mainly driven by the Earth's magnetic field. This region is called the **magnetosphere**. Charged particles can hardly leave this region because they are guided along the magnetic field lines they interact with. At the poles they are reflected back and forth and therefore held captive in the magnetic field lines they travel along. This phenomenon creates the hot, thin plasma which fills the magnetosphere. The magnetosphere itself is circumferentially flooded with the cold, dense plasma of the solar wind traveling at supersonic velocity. The solar wind is not able to penetrate the outer border of the magnetosphere, called the magnetopause, due to the Earth's strong magnetic field. It is forced to reduce its velocity at the so-called bow shock to a subsonic value and to stream around the magnetosphere. At this point, the solar wind with its high kinetic pressure compresses the magnetosphere on the day side of the Earth. By contrast, the magnetosphere on the night side of the Earth is formed into a tail extending beyond the orbit of the Moon.

How electron density varies with altitude is a consequence of the superposition of several single layers. Its origin [2.1.3] is explained qualitatively in Figure 2.1.6.

The density of the ionizable gas decreases with increasing height in the gravity field of the Earth. The intensity of the ionizable radiation, however, is highest at the top. As radiation is absorbed and used for ionization, the radiation intensity decreases in the downward direction. The **production rate of ions** is proportional to the quantities mentioned above and therefore has a strong maximum at medium altitudes. Because different gases exist within the air, different ionization layers of gas are created at different heights.

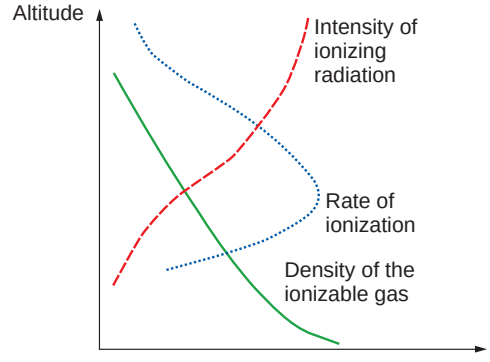


Figure 2.1.6: *Creation of an ionospheric layer (schematic).*

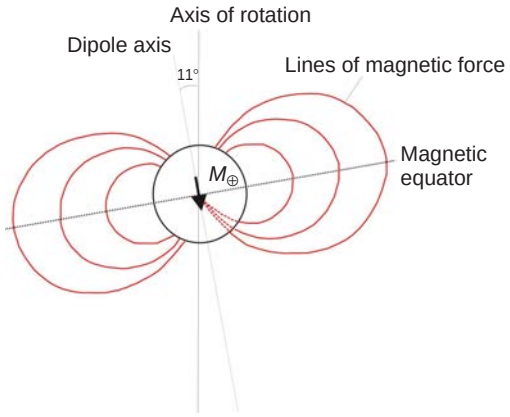


Figure 2.1.7: *Magnetic dipole of the Earth.*

The ionization of different molecule types is caused by different parts of the spectrum of the incoming radiation. The pattern mentioned above also applies to heat production, dissociation, and to all other processes where radiation acts on the air.

2.1.3.2 Earth's Magnetic Field

The magnetic field of the Earth can be approximated by a single dipole. Figure 2.1.7 schematically shows its structure. In this model the **magnetic dipole** moment is assumed to be approximately $7.7 \cdot 10^{22}$ A m^2. This dipole is tilted by 11° with respect to the rotational axis of the Earth.

If the Earth were located in space without any particles and without any electromagnetic field, the dipole structure of the magnetic field close to the planet would be preserved even far from the Earth. Based on what is known today, this is not the case. The magnetic field of the Earth is embedded in the particle stream of the Sun, called the solar wind. At the same time it overlaps with the interplanetary magnetic field. This interaction limits the magnetic field to a finite volume called the magnetosphere (see Figure 2.1.8).

On its side facing the Sun the magnetosphere is elliptical with the geocentric distance of the subsolar spot being 10 times the radius of the Earth (around 64 000 km).

Fluctuations of this distance up to several Earth radii can be observed. This variation depends on the physical properties of the interplanetary media and

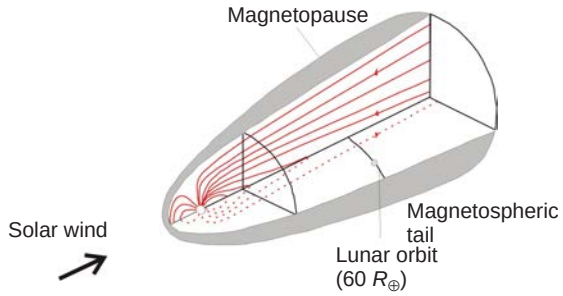


Figure 2.1.8: *Outer limits of the magnetosphere.*

especially on the pressure of the solar wind. On the side away from the Sun the magnetosphere is widely spread and forms a cylindrical body. This region is called the magnetospheric tail because of its similarity to the tail of a comet. The length of the tail is not exactly known and most likely quite variable. Usually, it significantly exceeds the lunar orbit (60 times the radius of the Earth) and its radius increases with the distance from the Earth, reaching values between 25 and 30 $R_\oplus$ at a distance of 200 $R_\oplus$.

2.1.3.3 Movement of Charged Particles within the Magnetosphere

Most particles populating the magnetosphere are ionized and therefore charged. Additional forces act on these charged particles moving inside the magnetic field. They are a function of the velocity of the charged particle. To describe this effect a balance of forces on a single particle is assumed. In this very simple model the following forces are taken into account: the inertial force F_T, the velocity-dependent magnetic field force F_B, and other external forces that are independent of velocity F_S. Friction forces and the repercussion on the magnetic field of the movement of the particles are not taken into account. The magnetic field is assumed to be a dipole. In this scenario we are looking for solutions of the differential equation

$$m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = \boldsymbol{F}_S + q\,(\boldsymbol{v}\times\boldsymbol{B}) \qquad (2.1.7)$$

where:
m = the mass of the particle,
t = time,
q = charge,

2

$\boldsymbol{v}$ = velocity
$\boldsymbol{B}$ = the surrounding magnetic field.

If we divide this differential equation into two components, one along the line of magnetic flux, the other one across it, we get

$$m \frac{\mathrm{d}v_{\text{parallel}}}{\mathrm{d}t} = \boldsymbol{F}_{S_{\text{parallel}}} \quad (2.1.8)$$

$$m \frac{\mathrm{d}v_{\text{vertical}}}{\mathrm{d}t} = \boldsymbol{F}_{S_{\text{vertical}}} + q\,(\boldsymbol{v} \times \boldsymbol{B}) \quad (2.1.9)$$

The first equation corresponds to the equation describing the "regular" movement of neutral gas particles. In case the external forces do not depend on time, the equation can be integrated directly:

$$\boldsymbol{V}_{\text{parallel}}(t) = \boldsymbol{V}_{\text{parallel}}(t_0) + \frac{\boldsymbol{F}_{S_{\text{parallel}}}}{m}(t - t_0) \quad (2.1.10)$$

The solution of Equation 2.1.9 is much more difficult and depends significantly on the configuration of the outer magnetic field. Therefore, it is helpful to break down the total movement into single components. In fact, this is often possible since these different single components are on different time scales and can therefore be treated independently. As a result, the following special cases occur, among others:

(1) $\boldsymbol{F}_{S\text{vertical}} = 0$, $\boldsymbol{B}$ = homogeneous → gyration
(2) $\boldsymbol{F}_{S\text{vertical}} = 0$, B gradient → oscillation
(3) $\boldsymbol{F}_{S\text{vertical}} = 0$, B gradient perpendicular to $\boldsymbol{B}$ → drift
(4) $\boldsymbol{F}_{S\text{vertical}} \neq 0$, $\boldsymbol{B}$ = homogeneous → drift.

A magnetic field perpendicular to the movement of the particles in response to gravity causes the charged particles to gyrate along the magnetic field lines. The magnetic field becomes stronger as it approaches the poles. Because the charged particles are forced out of an inhomogeneous magnetic field a **magnetic "bottle"** is created that traps the particles and reflects them from pole to pole (Figure 2.1.9).

The particles oscillate from one magnetic pole to the other and penetrate the atmosphere down to a height that is directly proportional to the strength of the magnetic field. Furthermore, there is an additional azimuthal drift of the charged magnetic particles due to the bending of the magnetic field lines. With this model of dipole approximation the reflecting points of all particle tracks move in circles of the same height. In reality the heights of the reflecting points fluctuate according to the local field strength. Particularly well known in this context is the so-called **South Atlantic anomaly** with its magnetic field strengths being significantly weak. With this anomaly, charged particles are able to penetrate extremely deep into the atmosphere.

2.1.3.4 The Radiation Belt (Van Allen Belt)

The highly energetic particle population of a radiation belt enclosed in the inner magnetosphere is a potential danger for spacecraft. High-energy particles are able to ionize especially electronic parts and therefore damage them. Such **ionization,** for example in an electronic semiconductor, is often reversible at first, but typically creates additional charge carriers leading to at least a temporary malfunction of the electronic semiconductor. If no countermeasures are taken this

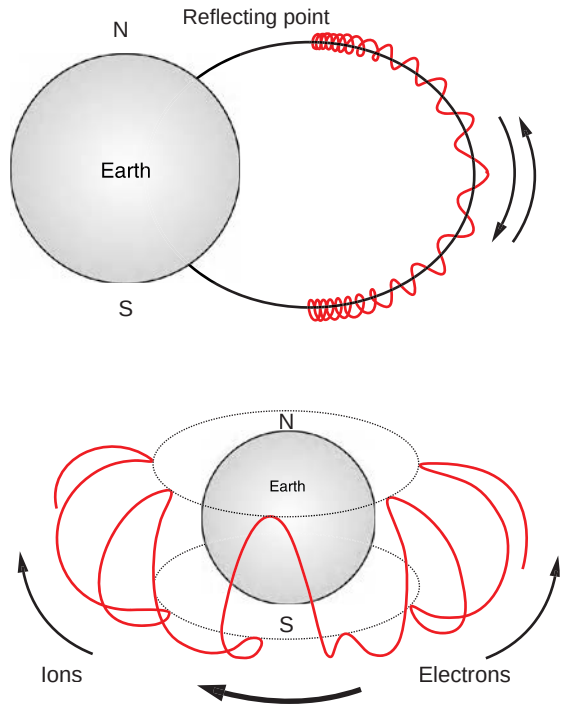


Figure 2.1.9: *Movement of charged particles in the magnetosphere.*

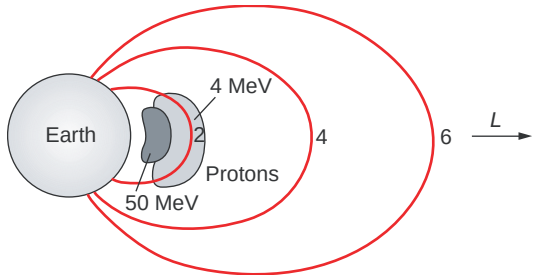


Figure 2.1.10: *Location of a radiation belt within the magnetosphere.*

may induce irreversible consequential damage that may even lead to a total loss of the spacecraft.

The location of maximum particle flow within the magnetosphere depends on the kind of particle, its energy and its magnetospheric condition. In the following the radiation belt is defined as the highly energetic component of the particle population with its lower border of 1 MeV for protons and 50 keV for electrons.

Figure 2.1.10 shows as an example the **distribution of the maximum general proton flow** of particles with energies of 4 or 50 MeV (according to [2.1.1]). For 4 MeV particles the maximum in the (magnetic) equatorial plane is at about $L = 1.8$ (at approximately 5000 km altitude), for 50 MeV particles it is at 3000 km (the inner radiation belt). The maximum equatorial flow is 10^{10} m^{-2} s^{-1} for 4 MeV particles and 10^{8} m^{-2} s^{-1} for 50 MeV particles.

Electrons by contrast have a maximum flow between $L = 3$ and 4 (the outer radiation belt). The physical properties of the particles of the higher energy radiation belts are relatively stable.

2.1.4 Effect on Spacecraft and Mission Design

On the one hand, it is self-explanatory that the mission and the spacecraft itself have to be designed with regard to the space environment to be encountered. On the other hand, mission and spacecraft designers try to benefit as much as possible from particular features of the space environment in terms of cost, energy, propellant, etc. This is why, compared to other environments, the space environment has such a huge impact on the design of a technical device. Disregarding some very particular conditions, the **space environment** can be described using the following categories:

- Gravitational fields
- Magnetic fields
- Electromagnetic radiation
- Atmospheres
- Energetic particle radiation
- High vacuum and plasma
- Contamination
- Rigid body interaction.

These conditions have to be evaluated and checked early in the design phase for their impact on the mission, because they contribute fundamentally to what decisions are made concerning orbit definition, mass budget, thermal control systems, and selection of components and materials. With the comprehensive experience collected since the beginning of the "space age," spacecraft designers can refer to a number of approved methods, models and data.

2.1.4.1 Gravity and Magnetism

The motion of a spacecraft exposed to the space environment is mainly determined by gravity. The orbit determination of every spacecraft in the gravitational field of a celestial body is based on Newton's law

$$F = -\frac{GM\,m}{r^2} \qquad (2.1.11)$$

where:

F = gravitational force,
G = universal gravitational constant,
M = mass of the central body,
m = mass of the spacecraft,
r = distance between the centers of gravity of both bodies.

However, this simple equation is only valid for:

- Spherical bodies in isolation from other masses
- The absence of external momentum (remaining atmosphere)
- Nonrelativistic conditions.

In the case of a nonspherical spacecraft, the gravity gradient will force the spacecraft axes to have the least moment of mass inertia pointing toward the central

body. Spacecraft having less strict requirements for attitude accuracy (> 1°) can actually make use of this natural effect for passive attitude control [2.1.6]. Conversely, active attitude control systems have to account for this effect.

The law of gravitation in its simple form is applicable only for orbits close to the central body where the influence of other celestial bodies on spacecraft motion is negligible. If this influence increases (e.g., for vehicles in geostationary orbits), a multibody problem has to be solved for orbit determination. Because of the mathematical complexity of these **multibody systems** they will not be discussed here; the reader is referred to publications dedicated to this topic.

More accurate orbit calculations which account for the deviation from the ideal spherical shape of central bodies require an exact model of their gravitational field. Meanwhile, a number of models of the Earth's gravitational potential have been derived using various methods, for example satellite tracking, Earth surface height measurements and surface-based gravity measurements. The basic principle of these methods is an infinite series of Legendre polynomials in latitudinal and longitudinal directions which solve Laplace's differential equation for a gravitational potential U in empty space: $\nabla^2 U = 0$.

The complete expression is given in the following formula:

$$U(r, \varphi, \lambda) = \frac{GM}{r}\left(1 + \sum_{n=2}^{\infty}\sum_{m=0}^{n}\left(\frac{\alpha}{r}\right)^n \cdot P_{nm} \cdot \left[C_{nm}\cos(m\ \lambda) + S_{nm}\sin(m\,\lambda)\right]\sin\varphi\right) \tag{2.1.12}$$

where:

α = reference axis of the ellipsoid,
n, m = degree and order of the harmonic terms,
λ, φ = geographical longitude and latitude,
C_{nm}, S_{nm} = spherical, harmonic coefficients,
P_{nm} = Legendre polynomial of the first order.

For example, the harmonic terms for $m = 0$ and $n = 2$ describe the Earth's polar flattening.

Nowadays numerical models of the **Earth's gravitational potential** are known to the degree and order of 70 and allow for an accuracy of orbit determination within a few centimeters. ESA's standard model currently in use is the JGM-2 (Joint Gravity Model 2) [2.1.7]. The models are steadily being improved, among other methods by satellite-based remote sensing measurements of the geoid (e.g., CHAMP, GRACE).

The increasingly detailed knowledge of gravitational fields in conjunction with high-power computing facilitates allow the use of gravitational fields for difficult orbital maneuvers, especially during interplanetary missions. Some sophisticated mission scenarios become feasible only with the gravitational help of one or more celestial bodies which can accelerate or decelerate a spacecraft during so-called fly-by maneuvers (i.e., flying by a celestial body to use its gravitational force for maneuvering). Fuel consumption can be reduced considerably using this approach. The most recent examples are the acceleration of the New Horizons Pluto probe during a Jupiter fly-by as well as the deceleration of the comet-chaser Rosetta during a Mars fly-by.

Within the combined gravitational field of the Sun and Earth as well as within all multibody systems a finite number of points exist in space where all acting gravitational forces are balanced (potential-free zones). These points are the so-called **Lagrange points**. They can be occupied by spacecraft using them as fixed observation points with respect to the two interacting celestial bodies. Referenced to the Earth–Sun system, Lagrange point L1 lies on a straight line crossing the centers of gravity of the Earth and Sun, 1.5 million kilometers before the Earth as seen from the Sun. L2 and L3 are located on the same line behind the Earth or beyond Earth's orbit 180° away from Earth's position, respectively. L4 and L5 are found on the Earth's track around the Sun at ±60° angular distance from the Earth. L1, L2 and L3 are metastable points whereas L4 and L5 are considered stable points. Spacecraft on Lagrange points follow either so-called halo orbits or Lissajous orbits around these points and permanently have to be corrected for orbital perturbations in order not to drift away. As opposed to the elliptic halo orbit, the Lissajous orbit requires fewer orbital correction maneuvers and thus also less fuel. It is not co-planar to the Earth–Sun plane but contains planar as well as vertical components. Figure 2.1.11 illustrates the distribution of the Lagrange points within the Earth–Sun system.

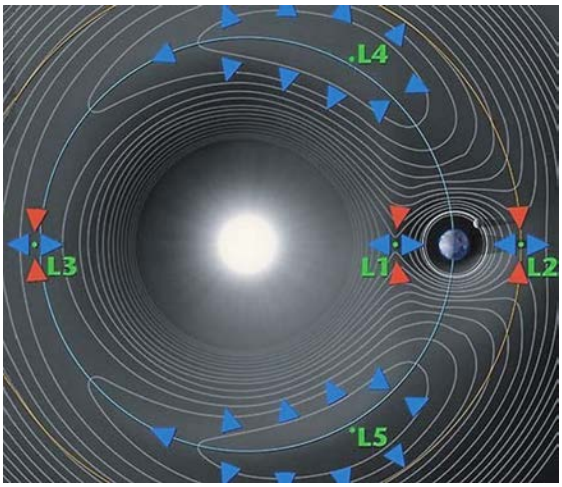


Figure 2.1.11: *The Lagrange points in the Earth–Sun system (Source: NASA).*

On a spacecraft traveling in space without any influence of either internal (propulsion) or external (atmospheric drag) forces, there is weightlessness as a result of the free fall within the gravitational field. The almost complete absence of gravity on-board has advantages but also implies special circumstances. One of the distinct advantages is the fact that mechanical devices and mechanisms can be designed for relatively small static and dynamic loads. The highest structural loads to be encountered are those arising during launch and reentry. One drawback of weightlessness on-board a spacecraft is the uncontrolled sloshing of liquids (e.g., propellants) inside the storage tanks. In order to avoid extensive sloshing and to assure the supply to the fuel pipes, the liquid has to be pre-tensioned. Pre-tensioning is achieved, for instance, in special tanks which make use of the liquid's own surface stress, by using membranes or by spinning.

Magnetism is a quite different physical phenomenon from gravitation. However, there are some similarities. The relevant magnetic fields for space flight are also bound to celestial objects. In contrast to gravitation, the origin of magnetic fields is related to magnetodynamical effects in the interior of the celestial body and is not just related to the existence of the body itself. Due to their dynamic origin, magnetic fields are not constant in their temporal or spatial dimensions, and, moreover, these fields are perturbed by charged particles or by plasma like the solar wind, which results altogether in more complex mathematical models.

Similar to the models of the geopotential there are also **magnetic models**. For the Earth, however, several models are required, depending on the radial distance and due to the strong interaction with the solar wind:

- Simple or eccentric dipole models, which may be applicable to LEO missions [2.1.8]. A standard reference model is the IGRF (International Geomagnetic Reference Field) model [2.1.9].
- The interaction of the solar wind with the geomagnetic field results in the deformation of its periphery: It is compressed in the direction of the Sun and builds up a tail in the direction opposite to the Sun [2.1.10].

With respect to space flight and the design of a spacecraft, magnetic fields imply the following interactions and consequences:

- The motion of a spacecraft within a magnetic field results in induced electromagnetic forces. Inside the spacecraft this may lead to electrical potentials in the range of several volts. Provided that the electric circuit is closed, electric current is generated inside the spacecraft and in the neighboring plasma. In so-called **tethered satellite missions** this effect might be used to generate electrical power for on-board power supply.
- In case a satellite has an **internal magnetic moment**, the external magnetic field exerts magnetic forces on the spacecraft. This effect may be used with smaller satellites to perform attitude control along the magnetic field.
- Magnetic fields generate radiation belts and the acceleration of charged particles in the direction of the magnetic poles causes certain orbits to be exposed to higher levels of radiation (see Section 2.1.4.4).

Accordingly, mission and spacecraft designers try to avoid the detrimental effect of magnetic fields as far as possible, other than for specific missions.

2.1.4.2 Electromagnetic Radiation

Electromagnetic radiation approaches a spacecraft from all directions (omnidirectional) with oscillating

2

intensities in time as well as in space. However, the impact on the operation of a spacecraft is mainly dominated by the electromagnetic radiation emitted by the Sun. The consequences of being exposed to this radiation are manifold and induce high demands on the design process.

The influences of electromagnetic radiation on a spacecraft in our Solar System can be categorized as follows.

Thermal Influence

The energy of the impacting electromagnetic radiation is mainly converted into **thermal energy**, which makes it the most important external parameter for the spacecraft's thermal balance. Relevant for the **thermal balance** are direct and reflected solar radiation, as well as the radiation from the Earth and from other celestial bodies (see Sections 2.1.2 and 2.1.3).

Therefore, in the design phase a detailed analysis of the chronological sequence of amplitudes, incident angles and wavelengths of the energy flux to be expected over the complete mission has to be performed. The information gained is used to design the **thermal control system** (TCS) of the spacecraft (see Section 4.3). Particularly challenging for the TCS are the extreme temperature gradients between the sides facing the Sun and those facing black space, or between Sun illumination phases and eclipses. Consequently, the linear expansion of the materials used is a critical design parameter. Interstellar electromagnetic radiation sources are not relevant for the thermal balance of a spacecraft. However, the temperature of the cosmic **background radiation** ($T = 2.7$ K) plays a major role in the design of specific science probes (IR telescopes) and their thermal balance since the operating temperature of their instruments is below the temperature of the background radiation.

Chemical Influence

Within the Solar System, the densities of the energy flows of short-wave solar electromagnetic radiation (UV, X-ray) are high enough to cause a change in the atomic structure of the radiated material. Electrons are ripped out of their atomic relation and move onward as free electrons. The consequence for a spacecraft is a number of mechanisms which can be useful on the one hand, but can also cause unwanted side effects on the other hand.

The effects on electrical characteristics are as follows.

Photoelectric Effect

Solar cells convert the free electrons into electric power by using the **photovoltaic effect** (see Section 4.2). Solar arrays are the basis of the power supply system of almost all of today's spacecraft. Today, they create an area-related power of approx. 350 W/m^2. Due to the larger distance to the Sun, solar arrays do not create sufficient power to supply a spacecraft beyond the inner planets. This is why radioisotopic thermal generators (RTGs) have to be used instead for missions to the outer planets.

Electrostatic Charging

An unwanted side effect of the removal of electrons from their atomic structure at the surface of the spacecraft is **electrostatic charging**. Positive charge is created on the illuminated side as opposed to the non-illuminated side (differential charging). This positive charge can partly compensate for the negative charging of the highly energetic electrons of the plasma ambiance (see Section 2.1.4.4). The charge level is determined by the attitude of the spacecraft relative to the Sun, among other factors. Satellites on Sun-synchronous polar orbits with one side permanently facing toward the Sun experience a higher risk than spin-stabilized interplanetary probes do, for example.

Due to solar flares the amount of short-wave radiation may increase a thousandfold or even more [2.1.4]. This may cause spontaneous discharge effects on the structure of the spacecraft. The change between solar illumination phases and eclipses may also trigger electric discharges. Moreover, operational events such as orbit maneuvers, turning on the downlink telemetry, or similar electronic activities, often are triggering electrostatic discharges [2.1.5]. During a discharge process high electric currents either flow on the satellite surface (in case of **differential charging**) or even discharge as an electric arc between the spacecraft and the surrounding plasma (**arcing**).

Possible consequences of these discharge processes are:

- **Degradation** of solar cells
- Degradation of optical sensors
- Acceleration of the degradation of the surface coatings.

A simple preventive measure is to use conductive materials for the outer structure of the spacecraft. In the case of solar arrays the use of such materials is, however, limited. As a compromise between electrical conductivity and optical transparency, indium oxide is used for the cover glass layer of the solar cells. Thus, the surface-related electrical resistance can be reduced to under 5000 Ω/cm^2 [2.1.4].

Change in Electrical Resistance
The removal of electrons from their atomic structure caused by UV radiation leads to only minor changes in the electrical resistance of electronic components.

Effects on the mechanical/optical characteristics include:

- **Embrittlement:** This is another kind of damage to the material caused by hard UV radiation. It has an effect on the stiffness of the mechanical structure and is particularly relevant at the component level for mechanisms or mechanically stressed components. Much polymer material tends to embrittle easily [2.1.4].
- **Darkening:** Due to a change at the atomic level, a darkening takes place in optically transparent material. This also includes optical lenses, cover glass layers of solar cells, and the glue between the cover glass and solar cell. Direct consequences are lower illumination of optical instruments and solar cells as well as an increase in the surface temperature of the solar cells due to the increase of the α/ε relation (see Section 4.3) of the cover glass layer, which is important for the heat balance.

Both the poorer illumination of the cells and the increase in cell temperature lead to a lower efficiency of power generation.

2.1.4.3 Atmospheric Influences

Naturally, atmospheric effects only become significant in close distance to planet surfaces. Consequently, the atmosphere of the Earth is only relevant for spacecraft operating in LEO. Although very thin at altitudes above 120 km, the **neutral gas atmosphere** has an effect on a spacecraft in LEO due to the **atmospheric drag** it causes.

As already mentioned in Section 2.1.1, a slipstream arises behind the spacecraft during the crossing of the residual atmosphere, meaning an area with significantly lower ambient pressure. This area can be used for very specific experiments that have high requirements with respect to the **vacuum**.

The composition and density of the residual atmosphere are strongly linked to **geomagnetic and solar incidents** since they heat up and therefore expand the upper atmosphere. At a height of 500 km, for example, the density of the atmosphere may vary by a factor of 100 between solar minima and maxima. Consequently, the atmospheric drag experienced by the spacecraft is subject to strong fluctuations which have to be taken into account for mission design and lifetime calculations.

One example of the significant effects of the oscillating atmospheric drag in LEO is the uncontrolled deorbiting of the science research module **Skylab** in 1979. A period of extremely high solar activity led to an unexpected increase of atmospheric drag, forcing the module to reenter the Earth's atmosphere even before a rescue mission could be initiated. The ISS loses between 100 and 200 m of height in orbit each day. This is why it has to be regularly reboosted into its nominal orbit either by means of its own thrusters or by those of a visiting, docked spacecraft. Publicly accessible indices help to predict the characteristics and the condition of the residual atmosphere. The atmospheric drag expected during a mission can be calculated; it is a function of the cross-sectional area of a spacecraft, its starting date, its orbit and its inclination.

The **atomic oxygen ratio** within the residual atmosphere creates an aggressive environment for the material of a spacecraft. This is not only the result of its chemical responsiveness, but also and more likely a combination of the latter with the high **relative velocity** of about 8 km/s with which the atoms hit the surface. The well-known effects are erosion, formation of stable oxides and chemoluminescent glow (spacecraft glowing). Due to **erosion**, volatile products are created, resulting in a recession of the surface and a self-contamination of the spacecraft. Highly susceptible materials are Kapton, polyethylene and pure silver [2.1.4], [2.1.15]. The large amount of silver that is used in solar arrays has to be therefore protected from direct oxygen flow.

Secondary effects of the erosion process are irreversible degradation of material properties (optical, thermal, mechanical, electrical), degradation in the quality of lubrication due to oxidation, and possible cracking within surface coatings. Spacecraft with high life expectancy (e.g., the ISS) and particularly sensitive optical instruments operating in an orbit below 500 km have to take into consideration the impact of atomic oxygen.

The short-wave radiation of the Sun causes ionization of parts of the upper atmospheric layer (photoionization, see Section 2.1.3). The density of the generated plasma within the ionosphere reaches its maximum at a height of 300 to 400 km. Nevertheless, the density of the neutral atmosphere below 1000 km is still considerably higher. The ionosphere has a significant influence on the diffusion of radio waves. Radio waves whose frequency is below the plasma frequency $f_p \approx 9000 \sqrt{n_e}$ (with n_e = density of electrons per cm^{-3}) are reflected at the ionosphere. The additional existence of the Earth's magnetic field leads to a reversion of the polarization of all electromagnetic radiation passing it [2.1.4].

These effects may have a negative influence on the communications with a spacecraft.

Apart from erosion caused by atomic oxygen, the atmospheres of other celestial bodies exert similar forces. It is worth mentioning the controlled deceleration of orbiters in a planet's atmosphere (**aerobraking**) in order to reach a desired orbit around that planet while simultaneously reducing fuel consumption significantly (Mars Global Surveyor, Venus Express).

The interaction of a spacecraft with the atmosphere during launch/landing and reentry is described in Section 2.3.

2.1.4.4 High-Energy Particle Radiation

During its mission every space vehicle is exposed to a continuous particle flux of varying intensity. High-energy particles with energies in the MeV range hit the spacecraft at high velocity and produce significant disturbances and danger. The density of the particle flux and its effects are considered early in the design phase. The radiation environment drives the selection of the operational orbit during the design phase. The expected radiation dose is determined by the utilization of dedicated models, and their effects are characterized depending on particle parameters. The following sources produce particles which influence space vehicles:

- Radiation belt
- Solar flares
- Cosmic particle radiation
- Secondary particle radiation
- Other radiation sources.

High-energy solar wind electrons and protons are captured by the Earth's magnetic field and concentrated in **radiation belts** (Van Allen belts). The inner radiation belt reaches closest to the surface of the Earth (about 500 km) in the South Atlantic area due to the Earth's magnetic field being tilted and shifted relative to its rotational axis. This is known as the **South Atlantic anomaly** (SAA). It dominates the doses of high-energy particles for vehicles in LEOs with low inclinations. This phenomenon needs to be taken into account during the spacecraft design phase. The radiation belts are crossed by LEOs as well as higher orbits (GEO and beyond). The effect of protons is more dominant in low orbits, whereas electrons have a larger influence in higher orbits. Due to the geometry of the Earth's magnetic field, spacecraft in polar orbits (> 80° inclination) are directly exposed to the solar wind. In these orbits the maximum tolerable **radiation dose (equivalent dose)** for astronauts is reached within about 5–7 days. Particles originating in the radiation belts cause degradation of electronic components, solar cells and materials; they are responsible for electric charging and so-called **single event phenomena**. In addition, they reduce the sensitivity of highly sensitive optical instruments (**CCDs**). Collision with highly energetic particles can cause **radioactive radiation** (activation). This leads to an increased **background noise** in sensitive measuring equipment and threatens biological systems.

The **degradation** of electronic components is caused by the ionizing effects of colliding particles. Solar cells degrade due to the displacement or bulk damage of atoms within their crystal structure caused by the nonionizing component of particle radiation. The materials used are sensitive to ionizing as well as nonionizing particle radiation.

Electrostatic charging of a spacecraft is caused either by collision and "sticking" of high-energy charged plasma particles (usually electrons) on the outer surface (**surface charging**) or by the accumulation of high-energy particles in dielectric materials (wire isolation, circuit boards) within the spacecraft (**deep dielectric charging**). The latter phenomenon in particular can have dramatic consequences as the resulting discharge effects can take place via electronic circuits.

Single event phenomena are caused by single collisions of very high-energy particles such as heavy ions, protons or neutrons originating from cosmic particle radiation or solar flares. Even protons from the inner radiation belt can cause single event phenomena by producing heavy energetic particles upon impact. These phenomena are a severe threat to electronic components. The degree of destructions is distinguished by the **linear energy transfer** (LET):

- Single event upset: temporary switch of the logical state of an integrated circuit causing a wrong command to, for example, the attitude control system.
- Single event latchup: static change of a circuit state which can only be recovered by a reset.
- Single event burnout: destruction of an integrated circuit.

In order to assure mission success, appropriate measures have to be taken in the design phase against single event phenomena [2.1.4]. Solar flares and **coronal mass ejections** (CMEs) produce an enormous number of high-energy particles which are propelled into interplanetary space. This results in short-term and highly fluctuating radiation doses which act upon a spacecraft.

Galactic cosmic rays are composed of high-energy protons (83%), alpha particles (13%), electrons (3%) and atomic nuclei with atomic numbers greater than 2 (1%). The energy of cosmic rays ranges between 100 and 10^{20} MeV [2.1.4]. Despite the low density of intergalactic particle radiation (in the order of a few particles per square centimeter per second) they do pose a threat to spacecraft due to their high impulse. Very high-energy protons can hit with almost the speed of light and thus cause destruction to sensitive materials, electronic components and biological tissue.

Secondary particle radiation is made up of particles which are generated by the impact of high-energy particles and then themselves interact with the spacecraft and cause disturbances.

Other sources of energetic particle radiation are ions generated by interaction with the upper atmosphere and neutrons as well as particle emissions from radioactive sources on-board a spacecraft (e.g., radioisotopic thermal generators (RTGs)).

2.1.4.5 High Vacuum

The very low ambient pressure in space beyond the neutral atmosphere (> 100 km altitude) also has a determining influence on the design and operation of a spacecraft. At an altitude of 500 km above the Earth's surface, for example, the **barometric pressure** is reduced to 10^{-7} Pa, whereas at the distance of a geostationary orbit (36 000 km) it has already converged to the value of interstellar space, approx. 10^{-15} Pa. A set of mechanisms and/or physical processes have to be highlighted in this context:

- Outgassing/sublimation
- Missing natural convection
- Change of material properties such as mechanical strength and life span/material fatigue
- Cold welding.

Due to the low ambient pressure, absorbed gas and water vapor escape from the materials. **Sublimation** describes the process of the evaporation of atoms/molecules from the surface as soon as the ambient pressure achieves and/or falls below the specific **steam pressure** of the material. The generic term **outgassing** refers to gases or particles escaping from the surface of a material. The outgassing rate increases with ambient temperature. Outgassing involves **mass loss** and a change in the surface properties of the materials concerned. Structural problems are not expected from outgassing. The following materials particularly tend to outgassing:

- Water
- Solvents
- Additives
- "Uncured monomeric material"
- Deposits which contaminated the spacecraft before or during the mission.

The outgassing products represent a danger for sensitive components, for example optical instruments, thermal coatings and high-voltage devices.

2

In the following section the influence of these outgassing products on the self-contamination of the spacecraft is described.

Traditional lubricants used on Earth are not suitable for space applications since they generally possess a high specific steam pressure and are often based on absorbed gases or water. As an alternative, coatings of graphite or molybdenum disulfite (MoS_2) are applied.

Heat Transfer

Due to the extremely low particle density of space, the process of convection, which is very important on Earth, plays a rather negligible role. The heat balancing between spacecraft and space environment takes place almost exclusively by radiant heat. The heat energy is exchanged internally by radiation and conduction. This is a major challenge for the design of the **temperature control system** (TCS), particularly for spacecraft which have to dissipate excessive heat.

Changes in Material Properties

Mechanical strength, life span and material fatigue are subject to change in the **high vacuum** environment. Generally, the life span of mechanically stressed construction units increases. For a lot of materials the fatigue behavior improves by more than one order of magnitude, although for some materials (e.g., nickel, Inconel 550) no statement can be given and reference is made to current materials research results. Material behavior in high vacuum is currently a subject of research on the ISS.

The increase in the mechanical strength of glass with falling ambient pressure is very well documented. It triples when the ambient pressure decreases from 1 bar to 1 mbar.

Cold Welding

Metallic parts lying close to each other can weld together due to the escape of the intermittent gas layer existing on Earth. Particularly with mobile parts this effect is regarded as critical.

2.1.4.6 Contamination

Contamination of Spacecraft

An individual forecast of the degree of contamination is very difficult to make since a wide range of different materials with very variable degassing characteristics is involved. Beyond this, the complexity of a contamination model is still further increased due to the interaction of the outgassing products with surfaces, the residual gas atmosphere, electromagnetic radiation, as well as high-energy particle radiation.

As sources of **contamination** the following phenomena have been identified:

- Degassing, including decay products.
- Particle plumes from the combustion processes of propulsion systems.
- Particle populations resulting from ion bombardment in plasma ("sputtering").
- Impact of micrometeorites and space debris.
- Ignition of pyrotechnic units or activation of release mechanisms.

Molecular degassing products partly disintegrate under the influence of short-wave solar radiation, high-energy particle radiation, atomic oxygen and electrostatic discharge processes. Thus, the contamination degree of a spacecraft is also influenced to a certain degree by solar activity and the interaction with the Earth's magnetic field; it therefore varies in time during the mission.

As a preventive measure against outgassing in space, a **bake-out** test is performed in a thermal vacuum chamber on the spacecraft before its launch (**baking**) in order to artificially start the outgassing process already on Earth.

So-called plumes are generated during the firing of the thrusters of the attitude and orbit control system and deposit on surfaces.

The impact of ions removes atoms from the surface and creates a particle population which may also distribute onto other surfaces (**sputtering**). Electrostatic charging may increase this effect as further ions are pushed toward the spacecraft due to the electric potential.

Consequently, the main focus concerning the effects of contamination is on the degradation of the efficiency of spacecraft systems or their subsystems due to:

- Particle contamination on delicate surfaces (mainly outgassing).
- Impact on thermo-optical properties.
- Impact on lubrication properties.
- Impact on electrical conductivity.

- Glowing, caused by atomic oxygen.
- Disturbances within the field of view of optical sensors.
- Light absorption.
- Scattering of light.
- Background noise in measurement devices.

Particle contamination on thermal surfaces has an impact on the emission and absorption behavior and therefore causes disturbances to which the temperature control system has to respond. Also the optical surfaces (lenses, mirrors) have to deal with this issue whereas, additionally, the transparency of the lenses and the reflecting power of the mirrors are adversely affected. In high-voltage components short circuits may occur due to outgassing. This is why, if possible, the commissioning of systems only starts after the outgassing process has reached a less critical level (after a few days, up to one month).

Contamination by the Spacecraft

When considering contamination, also the contamination of extraterrestrial matter by a terrestrial spacecraft should be taken into account.

One of the most dominant driving forces of humankind is to enhance its living space and its knowledge in search of the origin of its existence. This inevitably leads to contact between formerly biologically separated worlds (Apollo, Viking, SMART-1, Huygens, Rosetta).

There is a common agreement among spacefaring nations that unintended biological contamination of other celestial bodies by a terrestrial spacecraft has to be avoided. The scientific measurements undertaken in the search for extraterrestrial life must not be biased and the risk of destruction of possible life forms must be minimized.

Interplanetary probes and landers go through intensive cleaning and decontamination processes which are defined by standards. Similar rules apply for extraterrestrial material taken back to Earth from **sample return missions**. The current lunar initiatives of almost all space agencies (NASA, ESA, ISRO, CNSA), the exploration of Titan, Venus and Mercury (Cassini/Huygens, Venus Express, Bepi Colombo), and the already planned sample return or human space missions to Mars (USA, ESA), serve as references.

2.1.4.7 Micrometeoroids and Space Debris

Another dominating environmental effect on space vehicles is created by solid particles which are present in interplanetary space as well as within the Earth's orbit. These objects are either natural or artificial objects, also called **space debris**. Natural objects in Earth orbits are meteoroids, micrometeoroids, comets, asteroids and dust. Artificial objects include decommissioned satellites, rocket stages (partly exploded) and fragments, lost tools from space walks, paint particles and other such material.

The chances of a spacecraft being hit by a natural object are rather small. Exceptions are cyclically recurring meteoroid showers (Perseids, Leonids), whose intensities are higher than the annual average by a factor of 100. Artificial objects, however, especially small nontrackable objects, impose a much higher threat to human space flight and unmanned spacecraft because their number is very high. Objects more than 10 cm in size are tracked by radar stations, in contrast to smaller objects which are the most dangerous to spacecraft. Even very small particles in the submillimeter range can cause severe damage when they hit a spacecraft at speeds in excess of 10 km/s. Great effort is put into the development of multiple layer shielding mechanisms, which are based on the fragmentation of impacting objects and conversion of kinetic into thermal energy. The Ernst Mach Institute for Short-Time Dynamics, in Freiburg, Germany, has investigated this subject in detail by experiment and simulation.

A more detailed discussion of safety measures, collision probability and risks is given in Section 2.4.

Bibliography

[2.1.1] Prölss, G.W. *Physik des erdnahen Weltraums*, 2. Auflage. Berlin: Springer Verlag, 2004.

[2.1.2] ECSS E-10-04A. *The ECSS Space Environment Standard*. ESA/ESTEC, 2000.

[2.1.3] Kertz, W. *Einführung in die Geophysik II*, BI Hochschultaschenbücher, Bd, 535, 1969, unveränderter Nachdruck, 1985.

[2.1.4] Fortescue, P., Stark, J. *Spacecraft Systems Engineering*. Second Edition, Weinheim: Wiley-VCH, 2003.

[2.1.5] Houston, A., Rycroft, M. *Keys to Space*. International Space University, New York: McGraw-Hill, 1999.

[2.1.6] Wiesel, W.E. *Space Dynamics*. Hightstown, NJ: McGraw-Hill, 1989.

[2.1.7] Nerem, R.M. *et al.* Gravity Model Development for TOPEX/POSEIDON: Joint Gravity Models 1 and 2. *J. Geophys. Res.*, **99** (12), 24.421–24.447, 1994.

[2.1.8] Stern, D. Representation of Magnetic Fields in Space. *Rev. Geophys. Space Phys.*, **14**, 199, 1976.

[2.1.9] Peddie, N.W. International Geomagnetic Reference Field: The Third Generation. *J. Geomag. Geoelectr.*, **34**, 309–326, 1982.

[2.1.10] Stern, D.P., Tsyganenko, N.A. Uses and Limitations of the Tsyganenko Magnetic Field Models. *EOS, Trans. Am. Geophys. Union*, **73** (46), 489, 1992.

[2.1.11] Tribble, A.C. *The Space Environment: Implications for Spacecraft Design*. Princeton, NJ: Princeton University Press, 1995.

[2.1.12] Anderson, B.J., Smith, R.E. Natural Orbital Environmental Guidelines for Use in Aerospace Vehicle Development. *NASA Technical Memorandum 4527*, Marshall Space Flight Center (MSFC), Alabama, 1994.

[2.1.13] Ondoh, T., Marubashi, K. *Science of Space Environment*. Leipzig: IOS Press/LSL.de, 2001.

[2.1.14] Hastings, D. *Spacecraft: Environment, Interactions*, Cambridge Atmospheric and Space Science Series. Cambridge: Cambridge University Press, 2004.

[2.1.15] Hallmann, W., Ley, W. *Handbuch der Raumfahrttechnik*, 2. Auflage, Munich: Carl Hanser Verlag, 1999.

2.2 Orbital Mechanics

Oliver Montenbruck

Long before the launch of the first satellite, Sputnik, astronomers and those interested in celestial mechanics carefully studied the movements of the planets and moons. The physical models and methods of calculation they derived from these observations are also useful in many ways for describing the **motion of artificial Earth-orbiting satellites**. Their large number and closeness have conversely contributed to ever-improving understanding and more precise modeling of the forces involved. Having begun with Sputnik, which for the first time was able to prove that the Earth is actually pear shaped, modern satellites today like the GRACE formation allow us to follow seasonal variations in the Earth's gravitational field. An accurate description of the orbital mechanics of satellites is therefore not only important for operating the satellites themselves, but also of great relevance for geophysical research in general.

The following section begins by introducing the foundational concepts of the **laws of celestial mechanics** of satellites within the realm of Earth's gravity. They allow the modeling and **prediction of satellite orbits** and also form the basis for determining satellite orbits based upon radiometric or optical measurements. Besides the procedures for adjustment computation and filtering, various sensors and measurement methods used in spacecraft navigation will be introduced. The last section will use the example of remote sensing satellites and geostationary satellites to explain the methods of orbit determination and station keeping.

2.2.1 Orbit Modeling

2.2.1.1 Kepler Orbits

As described in the introduction, the motion of an Earth satellite follows the same basic physical laws that apply to the motion of planets around the Sun, or the Moon around the Earth. They were first derived by Johannes Kepler from observations of the planets. Later, with the help of Newton's law of universal gravitation, these motions could also be physically understood. Transposed to the case of a satellite in Earth's gravitational field, Kepler's three laws of planetary motion are:

1. The orbit of a satellite around the Earth is an ellipse (or more generally, a conic section), one focus of which coincides with the center of the Earth.
2. The radius vector from the Earth's center to the satellite sweeps over equal areas in equal time intervals.
3. The squares of the orbital periods of two satellites are proportional to the cubes of their average distance to the Earth's center.

Kepler's first law implies that the motion of a satellite takes place in an **invariable orbital plane** that passes through the Earth's center. The reason is that the Earth's attraction (to a first approximation) is always

directed at the Earth's center. Therefore, at no time is there a force perpendicular to the position vector $\boldsymbol{r}$ or velocity vector $\boldsymbol{v} = \dot{\boldsymbol{r}}$. Once an orbital plane has been established, it can no longer be left. Kepler's second law is also a direct result of these conditions and states nothing more than the conservation of angular momentum in a central force field.

By contrast, the elliptical orbit and the dependence of the orbital period on the size of the orbit are a result of the inverse-square reduction of the **gravitational acceleration**

$$\ddot{\boldsymbol{r}} = -\frac{GM_\oplus}{r^2} \cdot \frac{\boldsymbol{r}}{r} \qquad (2.2.1)$$

with the distance from the Earth's center (Newton's law of gravity). For Earth, the product of the gravitational constant and central mass has the value $GM_\oplus = 398\,600.4\ \mathrm{km^3/s^2}$.

If for simplicity's sake one examines a circular orbit, the equality of the gravitational and centrifugal forces directly results in Kepler's third law

$$a^3 n^2 = a^3 \left(\frac{2\pi}{T}\right)^2 = GM_\oplus \qquad (2.2.2)$$

which relates the orbital period T (or the angular velocity n) and the average orbital radius a of the satellites.

For low-altitude satellites (300–1000 km), **orbital periods** result ranging from 90 to 100 minutes, while a geostationary satellite at an altitude of 36 000 km requires exactly one day to complete an orbit (Figure 2.2.1).

Kepler's three laws make it possible to describe without additional tools a satellite's orbit and the time dependency of its motion. The form and size of an elliptical orbit are determined by the size of the semi-major and semi-minor axes (a and b) or the **eccentricity:**

$$e = \frac{\sqrt{a^2 - b^2}}{a} \qquad (2.2.3)$$

As shown in Figure 2.2.2, a denotes half the diameter of the ellipse along the line connecting the two foci (F_1, F_2). The semi-major axis therefore also represents the average of the smallest and greatest distance to Earth, when so-called perigee and apogee are reached. Perpendicular to the segment connecting

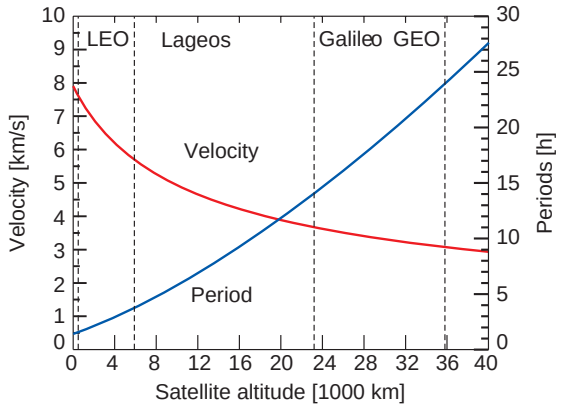


Figure 2.2.1: *Relationship between orbital altitude and orbital period for an Earth satellite.*

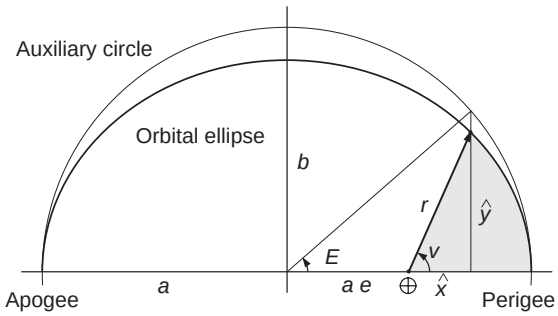


Figure 2.2.2: *Elliptical satellite orbit.*

both points (known as the line of apsides) the ellipse has its smallest axis, here with the size $2b$. The eccentricity is a descriptive measure for the deviation from an ideal circular form and expresses how far the foci are located from the middle point of the ellipse.

The distance of the satellite from the Earth's center varies throughout the course of an orbit between the extremes of $a(1 - e)$ during perigee and $a(1 + e)$ during apogee. For other points in the orbit the distance can be determined from the general conic section equation

$$r = \frac{a\,(1 - e^2)}{1 + e\cos(\upsilon)} \qquad (2.2.4)$$

This describes the relation between the orbital radius r and the true anomaly υ (the angular distance from perigee).

2

According to Kepler's second law, the **angular velocity** of the satellite in an elliptical orbit is not constant. The **true anomaly** $\nu(t)$ therefore varies periodically around the so-called mean anomaly, the angle M, that at a time t after passing perigee has the value $M = nt$. Only for circular orbits are the true anomaly and the mean anomaly identical at all times. By contrast, for elliptical orbits the angular velocity is greatest at perigee due to the proximity to Earth and the true anomaly at this point exceeds the mean anomaly. With increasing distance, the motion of the satellite is slowed down until at apogee both magnitudes once again match.

Beyond these general observations, using a combination of the conic section formula and Kepler's second law it is possible to derive the exact **time dependence of orbital motion**. Unfortunately, this reveals that a closed-form representation of the orbital position as a function of time is not possible and that instead an implicit description must be chosen. To this end an auxiliary number is used, the so-called **eccentric anomaly** E. As shown in Figure 2.2.2, the general relation

$$\begin{aligned} \hat{x} &= r \cdot \cos\nu &&= a \cdot (\cos E - e) \\ \hat{y} &= r \cdot \sin\nu &&= a \cdot \sqrt{1-e^2}\sin E \end{aligned} \qquad (2.2.5)$$

applies between the Cartesian coordinates $(\hat{x},\hat{y})$ and the eccentric anomaly. Alternatively, from Kepler's second law and the conic section formula the relationship

$$E - e\,\sin E = M = n \cdot t \qquad (2.2.6)$$

between the eccentric anomaly and the mean anomaly M can be derived.

This relationship, known as Kepler's equation, can, however, only be solved iteratively for E. However, it allows for any given time t (or mean anomaly M) the determination of the eccentric anomaly and hence the true anomaly (and therefore also the position of the satellite in its orbit).

As a "recipe" for solving Kepler's equation the classical **Newton's method** presents itself as a way to find a root of the function. Assuming a starting value $E_0 = M$, successively better values are found

$$E_{i+1} = E_i - \frac{E_i - e\,\sin\,E - M}{1 - e\cos E} \qquad (2.2.7)$$

until the solution no longer changes within the desired precision. In all practically relevant cases Newton's method provides a fast convergence while being simple to apply and understand. For highly elliptical orbits ($e > 0.8$) it is also possible to choose the starting value $E_0 = \pi$ to improve convergence.

After first determining the position of the satellite in its orbit $(\hat{x},\hat{y})$ it must then be represented in a global **frame of reference** using a suitable transformation. A natural choice for satellites orbiting the Earth is a coordinate system oriented on the equator. The z-axis then corresponds to the Earth's axis and the x-axis is aligned with the vernal equinox (Υ) that marks the direction to the Sun at the beginning of spring.

The orientation of the satellite's orbit relative to the equator and the vernal equinox is normally described using three angles whose meaning is illustrated in Figure 2.2.3:

- The **inclination** i measures the angle between the orbital plane and the equator.
- The **ascending node** is the point where the orbit crosses the equatorial plane from south to north and the intersecting line of the orbital and equatorial plane is the line of nodes.
- The **right ascension (or longitude) of the ascending node** Ω measures the angle between the vernal equinox and the ascending node.
- Finally, the **argument of perigee** ω describes the angle between the ascending node and the direction of perigee.

Based on these definitions the equatorial coordinates $\boldsymbol{r} = (x, y, z)$ of the satellite can be calculated through a

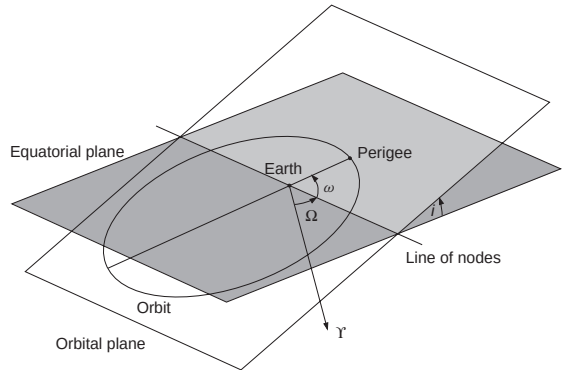


Figure 2.2.3: *Determination of the spatial position of the satellite's orbit using the three angles Ω, i and ω.*

series of three rotations from the previously determined coordinates $(\hat{x},\hat{y})$ in the orbital plane:

$$\boldsymbol{r} = \boldsymbol{R}_z(-\Omega)\,\boldsymbol{R}_x(-i)\,\boldsymbol{R}_z(-\omega)\cdot a\cdot\begin{pmatrix}\cos E - e\\ \sqrt{1-e^2}\sin E\\ 0\end{pmatrix} \tag{2.2.8}$$

The matrices

$$\boldsymbol{R}_x(\alpha) = \begin{pmatrix}1 & 0 & 0\\ 0 & +\cos\alpha & +\sin\alpha\\ 0 & -\sin\alpha & +\cos\alpha\end{pmatrix}$$

and

$$\boldsymbol{R}_z(\alpha) = \begin{pmatrix}+\cos\alpha & +\sin\alpha & 0\\ -\sin\alpha & +\cos\alpha & 0\\ 0 & 0 & 1\end{pmatrix}$$

describe here the elementary rotation around the x-axis and z-axis. Upon carrying out the corresponding calculations explicitly, the alternative formulation is found

$$\boldsymbol{r} = r\cdot\begin{pmatrix}\cos u\cos\Omega - \sin u\cos i\sin\Omega\\ \cos u\sin\Omega + \sin u\cos i\cos\Omega\\ \sin u\sin i\end{pmatrix},$$

in which $u = \omega + \upsilon$ (the "argument of latitude") describes the angle between the ascending node and the current orbital location.

For completeness, the relation for calculating the **velocity** of the satellite is given here:

$$\boldsymbol{v} = \boldsymbol{R}_z(-\Omega)\,\boldsymbol{R}_x(-i)\,\boldsymbol{R}_z(-\omega)\cdot\frac{\sqrt{GM_\oplus\, a}}{r}\cdot\begin{pmatrix}-\sin E\\ \sqrt{1-e^2}\cos E\\ 0\end{pmatrix} \tag{2.2.9}$$

A total of six independent **orbital elements**, also called **Keplerian elements**, are needed to uniquely describe the location and velocity of the satellite at any given point in time. Conversely, any given location and velocity vector can be associated with the six orbital elements [2.2.1], [2.2.2]. Ultimately, both representations are equivalent and have their own advantages and disadvantages, depending on the application. The meaning of each element is again explained in Table 2.2.1.

Table 2.2.1: *Keplerian elements.*

Orbital element		Meaning
a	Semi-major axis	Orbit shape, orbital period
e	Eccentricity	Orbit shape
i	Inclination	Spatial orientation
Ω	Right ascension of the ascending node	Spatial orientation
ω	Argument of perigee	Spatial orientation
M	Mean anomaly	Location in the orbit

2.2.1.2 Earth-Centered Orbits

The description of an orbit using the six Keplerian elements makes a simple and vivid description of the satellite's orbit in space possible. For many applications in Earth observation and communication, but also for satellite operation itself, it is primarily the motion of the satellite relative to the Earth's surface that is of interest. To provide this description, a coordinate system is used that is also oriented on the equator, the x-axis of which now points to the Greenwich meridian.

As illustrated in Figure 2.2.4, the absolute spatial (inertial) frame of reference and the Earth-centered (rotational) frame of reference differ by a rotation about the z-axis, reflecting the daily rotation of the Earth.

The angle Θ between the vernal equinox and the Greenwich meridian is commonly called **sidereal time** and is often expressed in time units (1 h represents 15°). Thus 24 hours of sidereal time (one sidereal day) represents exactly one complete rotation of the Earth around its axis. However, this period is approximately

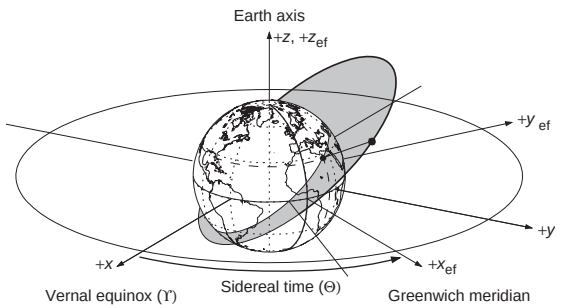


Figure 2.2.4: *Orbit of a satellite in the equatorial coordinate system.*

2

4 minutes shorter than one conventional day that orients itself on the Sun. The cause of this difference is the annual motion of the Earth around the Sun that causes the position of the Sun in the sky to shift daily by about 1°. For this reason, by the time the Sun once again reaches its highest point after one (solar) day, 4 minutes more have elapsed than the time necessary for one Earth rotation relative to the fixed stars.

For simple applications sidereal time as a function of **universal time** (UT1) can be computed using the formula

$$\Theta = 280.4606° + 360.985\,647\,3° \cdot d$$

in which d represents the number of days since noon on January 1, 2000. Strictly speaking, this formula (in a somewhat more complex form) describes the basis for the definition of universal time UT1 from the observation of the Earth's rotation. Due to irregularities in the Earth's rotation, the coordinated universal time for timekeeping (UTC) derived from atomic clocks is used. UT1 and UTC differ from one another by a maximum of 0.9 s and can therefore be used interchangeably for simple orbit predictions (visibility calculations, ground track representations, etc.) For higher accuracy requirements, the difference between these two time scales must be derived from publications of the International Earth Rotation Service [2.2.3].

Knowing the sidereal time, the Earth-centered position can be determined using a simple rotational transformation $\boldsymbol{r}_{\mathrm{ef}} = \boldsymbol{R}_{\mathrm{Z}}\,(\Theta) \cdot \boldsymbol{r}$ from which inertial coordinates can be determined. By finding the first derivative of the time, this expression is produced describing the change in time:

$$\begin{aligned} \boldsymbol{v}_{\mathrm{ef}} &= \boldsymbol{R}_z(\Theta) \cdot \frac{\mathrm{d}}{\mathrm{d}t}\, r + \frac{\mathrm{d}}{\mathrm{d}t}\, \boldsymbol{R}_z(\Theta) \cdot \boldsymbol{r} \\ &= \boldsymbol{R}_z(\Theta) \cdot \boldsymbol{v} - \begin{pmatrix} 0 \\ 0 \\ \omega_\oplus \end{pmatrix} \times \boldsymbol{r}_{\mathrm{ef}} \end{aligned} \qquad (2.2.10)$$

The velocity in the **rotating, Earth-centered system** therefore differs from the **inertial velocity** by a term which depends on the angular velocity

$$\omega_\oplus = 7.292\,12 \cdot 10^{-5}\ \mathrm{rad/s}$$

of the Earth and the distance of the satellite from the axis of rotation. For a satellite crossing the equator in low Earth orbit, the difference is about 500–600 m/s in an east–west direction. The ground path in the Earth-centered system therefore usually exhibits a somewhat different inclination to the equator than its (inertial) inclination would lead one to expect (Figure 2.2.5). Additionally, using the equation above, one can easily confirm that, as expected, the velocity of a geostationary satellite ($r = 42\,164$ km, $v = 3.075$ km/s) disappears in an Earth-centered system.

For simplicity's sake, in the discussion above the transition from an absolute spatial to an Earth-centered frame of reference only considered the daily rotation of the Earth. More precise calculations must additionally take into account that the rotational axis of the Earth is not fixed in space. Due to the torques of the Sun and Moon that work to right the Earth's axis, the axis precesses in a conical motion around the pole of the ecliptic. This precession has as period of 26 000 years. Superimposed on this are short-term variations known as **nutation** that lead to deflections of approximately 20″ from the mean orientation. The rotational axis also varies relative to the Earth's crust itself due to the polar motion of approximately 10 meters.

In an effort to create a uniform frame of reference despite these variations, different organizations created the **International Celestial Reference System** (ICRS) and the **International Terrestrial Reference System** (ITRS). The ICRS is oriented on the position of the equator and vernal equinox at the beginning of the year 2000. Today it is determined by radio astronomical observations of distant galaxies. The Earth-centered



Figure 2.2.5: *Ground track of a satellite in low Earth orbit. At 14 to 15 orbits per day each successive equatorial crossing is shifted by approximately 25° geographical longitude.*

ITRS is realized through a network of ground stations with GPS receivers and telescopes for laser distance measurements. The transition between the two systems is described by a series of transformations (precession, nutation, Earth rotation, polar motion) that cannot be covered here due to limited space. Interested readers are hereby referred to appropriate textbooks (e.g., [2.2.1], [2.2.2]) or monographs (e.g., [2.2.4], [2.2.5]).

2.2.1.3 Orbital Perturbations

The assumption of a point mass (or a spherical and homogeneous central mass) is the basis for the description of a satellite's orbit using Kepler's laws. For the Earth this assumption is certainly appropriate in a first approximation, but closer examination reveals a number of variations from this ideal picture. First is the fact that due to its rotation the Earth's radius at the poles is about 20 km smaller than at the equator. This **flattening (oblateness)** and the resulting deviation from the perfect hypothetical central force leads, among other things, to a continual change in the orbital planes of satellites. Other perturbations result from the Earth's atmosphere, which causes a continual braking of low-flying satellites and so a loss of altitude. For higher flying satellites the **gravitational perturbations** caused by the Sun and the Moon play a role. For exact orbital calculations, many effects (the inhomogeneous distribution of mass within the Earth, radiation pressure, tides, relativistic effects, etc.) must be accounted for. The magnitudes of the most important perturbations are illustrated in Figure 2.2.6.

Gravity

For a practical description of Earth's gravity, an expansion of the gravitational potential in terms of spherical harmonics can be employed. In general the potential of a point at the geocentric latitude λ and longitude φ at a distance r from the Earth's center has the form

$$V = \frac{GM_\oplus}{r} \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{R_\oplus^n}{r^n} P_{nm}(\sin\varphi) \cdot \left[C_{nm} \cos m\lambda + S_{nm} \sin m\lambda\right] \qquad (2.2.11)$$

Here $R_\oplus$ stands for the equatorial radius of the Earth, while P_{nm} stands for the associated Legendre polynomial of order n and degree m. The harmonic coefficients C_{nm} and S_{nm} of the gravitational field can be determined from observations of satellite orbits and supplemented by measurements of gravitational acceleration close to the ground. Where earlier models based on satellite observations were limited to lower dimensions (e.g., 30×30), today missions like CHAMP and GRACE provide much higher resolution **gravitational field models** (e.g., GGM02 with coefficients up to 160×160 [2.2.7]).

For a given potential $V(\boldsymbol{r})$, one obtains the corresponding acceleration from the gradient

$$\ddot{\boldsymbol{r}} = \nabla V(\boldsymbol{r}) = \left(\frac{\partial V}{\partial x}; \frac{\partial V}{\partial y}; \frac{\partial V}{\partial z}\right) \qquad (2.2.12)$$

that can be most easily calculated with the help of the recurrence relations described in [2.2.8]. For the main term $C_{00} = 1$ one thus gets Newton's law of universal gravitation describing the attraction of a spherical body with isotropic mass distribution.

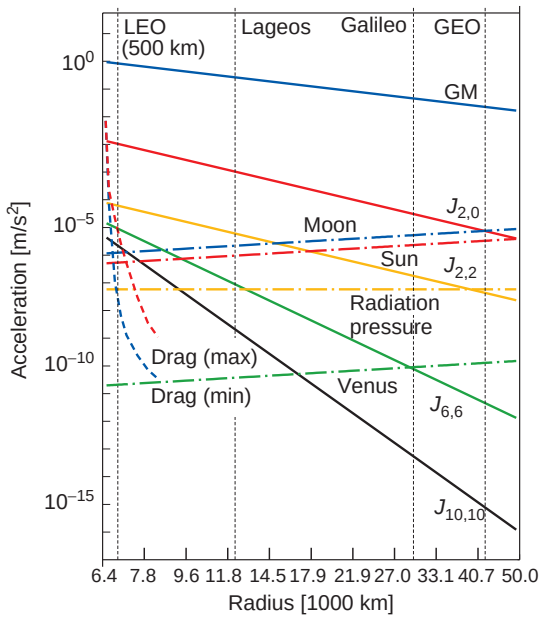


Figure 2.2.6: *Comparison of the perturbing accelerations affecting satellites at various altitudes. Indicated are the effects of Earth's gravitation (central term GM and harmonic term $J_{n,n}$), the perturbation from point masses (Moon, Sun, Venus) as well as the influence of radiation pressure and air resistance (drag).*

The Earth's **flattening** is reflected in the leading zonal coefficient $C_{20} = -J_2 = -1.082 \cdot 10^{-3}$, leading to a perturbation of the magnitude

$$\begin{pmatrix} \Delta\ddot{x} \\ \Delta\ddot{y} \\ \Delta\ddot{z} \end{pmatrix} = -\frac{3}{2} J_2 \frac{R_\oplus^2}{r^5} \cdot \begin{pmatrix} x - 5xz^2/r^2 \\ y - 5yz^2/r^2 \\ 3z - 5z^3/r^2 \end{pmatrix} \qquad (2.2.13)$$

In the mean over an orbit, a net torque results, to which the satellite responds by a slow rotation of the orbital plane. Depending upon the orbit altitude and inclination, this alters the right ascension of the ascending node with the rate

$$\Delta\dot{\Omega} = -\frac{3}{2} J_2 n \frac{R_\oplus^2}{a^2 (1-e^2)^2} \cos(i) \qquad (2.2.14)$$

that can range up to 8° per day. Similar variations

$$\Delta\dot{\omega} = +\frac{3}{4} J_2 n \frac{R_\oplus^2}{a^2 (1-e^2)^2} \left[4 - 5 \sin^2(i) \right] \qquad (2.2.15)$$

and

$$\Delta\dot{M}_0 = +\frac{3}{4} J_2 n \frac{R_\oplus^2}{a^2 (1-e^2)^{3/2}} \left[2 - 3 \sin^2(i) \right] \qquad (2.2.16)$$

can also be found in the argument of perigee and the mean anomaly. Additionally, all six orbital elements are subject to periodic perturbations, the amplitudes of which have an approximate magnitude of aJ_2, or about 10 km.

For precise orbital predictions many additional **gravitational field coefficients** besides J_2 must be considered to correctly describe the gravitational attraction of the Earth. It must also be considered that the contribution of terms of degree n to the acceleration diminishes by a power of $(n + 2)$ of the distance from the Earth's center. For satellites at middle and high altitudes (e.g., navigation satellites and geostationary satellites) field coefficients up to degree and order 10 are therefore usually sufficient for precise orbit calculations. By contrast, for lower altitudes (400 km) and geodetic applications, terms of up to 100 × 100 must be considered.

Gravity from Sun and Moon

Besides the Earth, other celestial bodies – most importantly the Sun and Moon – exert an acceleration on satellites which stems from the attraction between two masses. However, this attractive force affects not only satellites but also the Earth's center of mass. This means that only the difference between the two acceleration functions acts as an effective perturbing acceleration.

If one designates the mass and the geocentric position vector of the perturbing body as M_s and $\boldsymbol{r}_s$, the perturbing acceleration is

$$\Delta\ddot{\boldsymbol{r}} = +GM_s \frac{\boldsymbol{r}_s - \boldsymbol{r}}{|\boldsymbol{r}_s - \boldsymbol{r}|^3} - GM_s \frac{\boldsymbol{r}_s}{|\boldsymbol{r}_s|^3} \qquad (2.2.17)$$

For this purpose the perturbing mass is assumed to be a point, which for the practical calculation of periodic satellite orbits around the Earth is a fully acceptable approximation.

If the Earth, the satellite and the perturbing body are in a line, to a first approximation a perturbing acceleration of magnitude

$$\Delta\ddot{\boldsymbol{r}} \approx +2\,GM_s \frac{\boldsymbol{r}}{|\boldsymbol{r}_s|^3} \qquad (2.2.18)$$

results, directed away from the Earth. If the position vector of the satellite is perpendicular to the direction of the perturbing body, the resulting acceleration is only half as large and directed inward:

$$\Delta\ddot{\boldsymbol{r}} \approx -GM_s \frac{\boldsymbol{r}}{|\boldsymbol{r}_s|^3} \qquad (2.2.19)$$

This is illustrated in Figure 2.2.7.

If one compares the force exerted by the Sun ($M_{Sun} = 330\,000 M_\oplus$, $r_{Sun} = 150 \cdot 10^6$ km) and Moon ($M_{Moon} = 1/81 M_\oplus$, $r_{Moon} = 400\,000$ km), it can be seen that despite the Moon's much smaller mass, the effect exerted by it is about twice as large. The equation above also shows that the perturbing acceleration

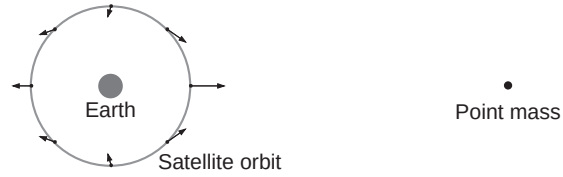


Figure 2.2.7: *Perturbing acceleration of a satellite by a point mass far away.*

grows linearly with the distance of the satellite from the Earth. This is particularly important to consider when modeling geostationary satellite orbits.

Air Drag

Besides the flattening of the Earth, for low-flying satellites atmospheric braking represents the greatest source of perturbation. In contrast to gravitational perturbations, here the effective acceleration is dependent upon the mass m of the satellite and its cross-sectional area A. For a given density ρ of the atmosphere, the drag can be modeled in the simplest case using the relationship

$$\Delta\ddot{\boldsymbol{r}} = -\frac{1}{2} C_D \frac{A}{m} \rho \cdot \boldsymbol{v}_{rel} \boldsymbol{v}_{rel} \qquad (2.2.20)$$

This is always related to the velocity vector $\boldsymbol{v}_{rel}$ relative to the direction of the atmosphere and grows as the square of the velocity.

The **drag coefficient** C_D represents the individual aerodynamic characteristics of a particular body. Usually it must be calibrated for a satellite when calculating the orbit. Values between 2.0 and 2.3 are often used as a rough estimate. Alternatively, the drag coefficient can be determined using model calculations that take into account the atmospheric conditions in the upper atmosphere [2.2.9]. However, these so-called **computational fluid dynamic calculations** (CFD calculations) are very complex and can usually only be employed for practical calculations after simplified macromodels are derived.

Figure 2.2.8 shows an overview of density variations in the Earth's atmosphere. As can be seen clearly, the density decreases rapidly with increasing altitude and the atmosphere can be largely disregarded for orbits above 1000 km. In contrast, for altitudes below 250 km the drag is so high that without regular maneuvers to raise the orbit, a rapid reentry of the satellite would be unavoidable. In general, because of the desire for a long life span, only a very few satellites are operated below 400 km.

In total, atmospheric drag results in a continual **reduction of orbital energy** and therefore of the semi-major axis. For a circular orbit this results in a continuous decrease in orbital altitude that at first proceeds slowly, then evermore quickly (see Figure 2.2.9). For strongly elliptical orbits the braking

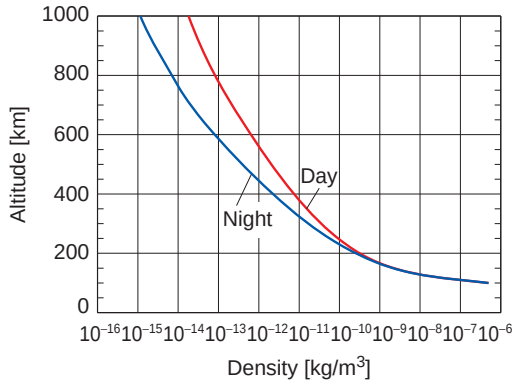


Figure 2.2.8: *Density variation of the atmosphere as a function of altitude for mean solar activity.*

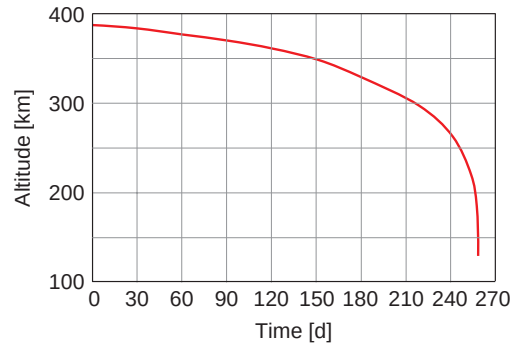


Figure 2.2.9: *Decrease in orbital altitude of Starshine 1.*

effect occurs primarily close to perigee. At the lowest point in the orbit the satellite dips into the lowest and therefore thickest layers of the atmosphere. At the same time the velocity is highest at this point, further increasing the effects of drag. As a result of this uneven braking the apogee altitude at first decreases, until the orbit is nearly circularized.

For a description of **atmospheric density** various **models** are available that offer a typical prediction accuracy between 10 and 30% [2.2.1], [2.2.2], [2.2.10]. Best known are the models of the Jacchia series (Jacchia-70 and further developments), the family of the "Mass Spectrometer and Incoherent Scatter" models (MSIS-86, NRLMSIS) and the "Drag Temperature Models" (DTM-94, DTM-2000). Besides the expected accuracy, when choosing an appropriate model computational costs are usually a concern. Models like Jacchia-70 that restrict themselves to

2

the total atmospheric density are usually more advantageous here than models that provide a detailed description of chemical composition.

Considering the existing uncertainties in density modeling and the poor predictability of the drag coefficient, a free scale factor must generally be estimated to compensate for the described errors when determining an orbit. Simple density models are usually sufficient for a good reconstruction of the orbit from observations. Conversely, more modeling effort is required to exactly determine the orbit for a pure orbit prediction.

Atmospheric density displays a pronounced **day–night variation** due to solar warming, with the maximum effect being reached about 2 hours after midday. Furthermore, it has been shown that atmospheric density is subject to large variations which correlate with solar activity (extreme ultraviolet rays) and geomagnetic activity. For this reason all density models require the inclusion of a value for solar flux (radio flux $F_{10.7}$ at a wavelength of 10.7 cm as an indirect measure for the UV flux) as well as the geomagnetic index (K_p). Both quantities must be determined by measurements and have only limited predictability. Long-term forecasting of satellite orbits in the atmosphere is therefore severely restricted.

Radiation Pressure

A second nongravitational force that must be considered when modeling satellite orbits results from the **radiation pressure of the Sun**. When photons with energy E strike the surface of a satellite and are absorbed, an impulse is transferred with a magnitude $p = E/c$. Close to the Earth, meaning within a distance of 1 AU (149.6 million km) from the Sun, the solar flux Φ has a seasonal average of about 1371 W/m^2. From the value of this so-called solar constant and the speed of light c, a pressure of

$$P_0 = \frac{\Phi}{c} = 4.57 \cdot 10^{-6}\ \mathrm{N/m^2} \qquad (2.2.21)$$

results.

If the light is not absorbed, but instead completely reflected, the impulse transferred and therefore the resulting light pressure is actually twice as high. For a Sun-facing cross-sectional area A and a mass m for a satellite, the magnitude of the acceleration is

$$\Delta\ddot{\boldsymbol{r}} = -(1+\varepsilon)\frac{A}{m} \cdot P_0 \cdot \left(\frac{1\ \mathrm{AU}}{r_{\mathrm{Sun}}}\right)^2 \cdot \frac{\boldsymbol{r}_{\mathrm{Sun}}}{r_{\mathrm{Sun}}} \qquad (2.2.22)$$

which always acts in a direction opposite to the Sun's vector. The reflectivity ε describes the relation between reflected and incidental sunlight. In its place the so-called **radiation pressure coefficient** $C_R = 1 + \varepsilon$ is often used, and as with the drag coefficient can be calibrated within limits. As a first approximation a value of 1.3 is often used.

As expected, radiation pressure is a particular concern for satellites that possess large solar panels for producing energy. For geostationary satellites with an area of up to 50 m^2, radiation pressure represents one of the greatest perturbing factors that must be considered when maintaining an orbit. This leads to a yearly variation in the eccentricity that must be compensated with regular maneuvers to prevent leaving the assigned geostationary window (see Section 2.2.3.3).

Precise modeling of radiation pressure is also important for navigation satellites or remote sensing satellites with tight requirements for knowledge of the orbit. Here part of the work to correctly model radiation pressure is done with optical ray tracing simulations and finite element models [2.2.9], [2.2.11]. In this way specific characteristics (absorption, reflection, thermal radiation, etc.) for individual components and materials can be accurately accounted for.

As a part of the radiation pressure calculation one must finally consider if or to what extent the satellite will be exposed to the Sun. In the simplest case a simple **cylindrical shadow model** is sufficient (Figure 2.2.10). The satellite is then not illuminated when it is on the night side of the Earth ($\boldsymbol{r} \cdot \boldsymbol{e}_{\mathrm{Sun}} < 0$) and its distance

$$d = \left|\boldsymbol{r} - (\boldsymbol{r} \cdot \boldsymbol{e}_{\mathrm{Sun}})\, \boldsymbol{e}_{\mathrm{Sun}}\right| \qquad (2.2.23)$$

from the shadow axis is smaller than the Earth's radius $R_\oplus$. The term $\boldsymbol{e}_{\mathrm{Sun}}$ describes the unit vector

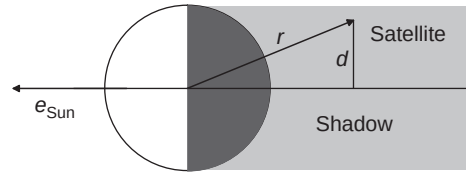


Figure 2.2.10: *Simple cylindrical model of the Earth's shadow.*

from the Earth to the Sun. Beyond the consideration above, more exact shadow models take into account the flattening of the Earth and the finite diameter of the Sun, which results in a separation into penumbra and umbra [2.2.2].

2.2.1.4 Analytical Orbit Models

With the help of perturbation calculations, the effects on satellite orbits of the forces discussed in the previous section can be captured analytically. Analogous to Kepler's orbit model, the models thus derived allow a direct, analytical **description of the orbital path** as a function of the orbital elements. The perturbation calculation can at once be used to calculate orbits, while also providing insights into the effect of individual components of the perturbation.

The heyday of analytical orbit models falls between the years 1950 and 1970, when computers were strictly limited in their availability. Due to their complexity and limited precision, analytical orbit models now play a significantly smaller role. They are, however, advantageous compared to theoretical investigations when a satellite's orbit is needed for an arbitrary epoch. In contrast to the numerical procedure described later, analytical models also allow the determination of a satellite's position in the distant past or in future epochs without intermediate steps.

Variation of the Elements

Kepler's orbit model discussed above describes the motion of a satellite in a central force field, the strength of which decreases as the square of the distance from the center. Orbital elements like semi-major axis, orbital inclination and mean anomaly at the epoch are constant quantities that characterize the orbit for all times.

Analytical orbit models utilize the consideration that small variations from an ideal $1/r^2$ force field must therefore be reflected in small temporal variations of the orbital elements. How these variations look in detail is described, for example, by the **Gaussian variational equations**. If one decomposes the perturbing acceleration $\boldsymbol{a}$ into three orthogonal components along the radial direction (R), perpendicular to the radial direction in the direction of motion (T) and in the direction of the orbit normal (N), this produces the following expressions for the **temporal variation of the orbital elements** [2.2.1], [2.2.13]:

$$\frac{da}{dt} = \frac{2}{n\sqrt{1-e^2}}\left[e\sin\nu \cdot a_R + \frac{p}{r}\cdot a_T\right] \tag{2.2.24}$$

$$\frac{de}{dt} = \frac{\sqrt{1-e^2}}{n\,a}\left[\sin\nu \cdot a_R + \quad (\cos E + \cos\nu)\cdot a_T\right] \tag{2.2.25}$$

$$\frac{di}{dt} = \frac{1}{n\,a^2\sqrt{1-e^2}}\, r\cos u \cdot a_N \tag{2.2.26}$$

$$\frac{d\Omega}{dt} = \frac{1}{n\,a^2\sqrt{1-e^2}}\,\frac{r\sin u}{\sin i}\cdot a_N \tag{2.2.27}$$

$$\frac{d\omega}{dt} = \frac{\sqrt{1-e^2}}{n\,a\,e}\left[-\cos\nu \cdot a_R + \left(1+\frac{r}{p}\right)\sin\nu \cdot a_T\right] - \cos i \cdot \frac{d\Omega}{dt} \tag{2.2.28}$$

$$\frac{dM_0}{dt} = \frac{1}{n\,a^2\,e} \cdot \left[(p\cos\nu - 2\,e\,r)\cdot a_R - (p+r)\sin\nu \cdot a_T\right] \tag{2.2.29}$$

If one first assumes an unperturbed orbit, it is then possible to calculate the perturbations along the orbit and – with corresponding effort – represent them as a periodic series expansion. This can then be integrated to achieve an analytical description of the change in the orbital elements over time. When necessary, the thus corrected orbital representation can again be inserted into the perturbation equations to obtain higher order perturbation theories.

A well-known example is the **perturbation theory** constructed by **W. M. Kaula** [2.2.14], which describes the changes to an orbit under the influence of the Earth's gravitational field. The perturbations can be divided into periodic and secular (growing over time) changes to the orbital elements (Figure 2.2.11).

Variations with a typical time scale of a single orbit are referred to as short-periodic perturbations. Long-periodic perturbations reflect the changes to the perigee altitude under the influence of the Earth's

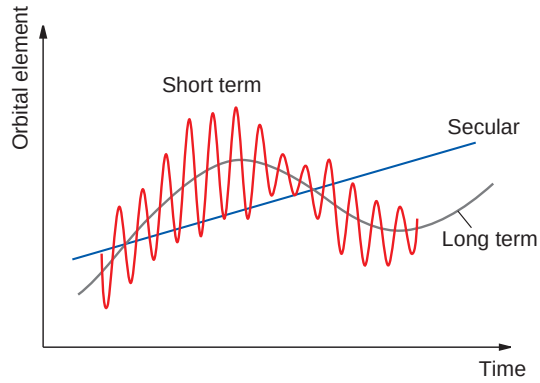


Figure 2.2.11: *Development over time of perturbations in the Keplerian orbital elements.*

flattening (see Section 2.2.1.3) or the monthly and yearly contributions of the gravitational perturbation from the Sun and Moon. Finally, secular perturbations describe the long-term changes that remain after averaging over the previously described time scales. Kaula's theory shows among other things that the secular perturbations from Earth's gravitational field only appear in the angular elements of the orbit. Perturbations of the semi-major axis and the eccentricity are, however, only subject to the periodical portion.

The SGP4 Model

Certainly one of the most popular orbit models is the "Simplified General Perturbations Model" No. 4 (SGP4) [2.2.15], [2.2.16]. It was originally developed for the needs of the North American Aerospace Defense Command (NORAD), an organization also responsible for the continuous monitoring of near-Earth objects. Part of this responsibility includes the regular determination of the orbital elements of all cataloged satellites, which within limits are also available to civilian users. Amateur radio operators in particular profit from this service, which they can use to easily generate current visibility predictions for their satellites.

The **NORAD orbital elements** are published in a two-line data format, which has earned them the nickname "two-line elements" (TLEs). The meaning of the individual fields in the TLE data is described in Figure 2.2.12.

When using the NORAD two-line elements it must be borne in mind that they represent "average" orbital elements. In contrast to "osculating" elements that at a particular time reflect the actual position and velocity of the satellite, average elements are free of model-specific periodic perturbations. They therefore show a much smoother progression than the osculating elements. Conversely, the average elements should never be used with a simple Keplerian orbit model, but rather only with the analytical orbit model that was used to generate them.

The SGP4 model employed by NORAD for this purpose was created around 1970 and is based upon an analytical orbit model from Lane and Crawford. This model combines modeling of the gravitational perturbations from Brouwer [2.2.17] with an analytical model describing drag. The atmospheric density at an altitude h is described using a power law of the form

$$\rho = 2.461 \cdot 10^{-8}\ \mathrm{kg/m^3} \left(\frac{42\ \mathrm{km}}{h - 78\ \mathrm{km}} \right)^4 \qquad (2.2.30)$$

For operational use the complete theory from Lane and Hoots was simplified and combined with the still common form of the SGP4 model. The secular and periodical perturbations are accounted for by the gravitational field coefficients J_2, J_3 and J_4. Strictly speaking, the representation of the periodical perturbations only applies to near-circular orbits.

In addition to the SGP4 model NORAD uses an extended form of the SGP4 model called the SDP4 model for high-altitude satellites. This considers additional perturbations caused by the gravity of the Sun and Moon, as well as **resonance terms** that are significant for orbits with periods between 12 h (GPS) and 24 h (geostationary satellites). The SDP4 model is used for the generation of two-line elements when the orbital period of a satellite is greater than 225 minutes (so-called "deep-space" orbits). This value corresponds to an orbital altitude of about 6000 km and serves the users of the two-line elements as a discriminating criterion for the choice of the correct orbit model.

The accuracy of the SGP4 orbit model and the two-line elements produced by it are illustrated in Figure 2.2.13 for a low-flying satellite (altitude 400 km). Close to the epoch the actual orbit is given within an accuracy of 1–2 km. This value reflects the magnitude of a series of short-period perturbations that are ignored in the SGP4 model. This then represents a

```
.........|.........|.........|.........|.........|.........|.........
1 25544U 98067A   06287.64456019  .00008182  00000-0  53355-4 0  8609
2 25544 051.6354 292.0281 0013277 096.0881 057.9543 15.76874518452008
.........|.........|.........|.........|.........|.........|.........
123456789012345678901234567890123456789012345678901234567890123456789
```

Row	Column	Description	Example	Meaning
1	1	Line number identification (= 1)	1	First line
	3–7	NORAD catalog number (Example 16609)	25554	ISS (Zarya Module)
	8	Security classification	U	Not classified
	10–17	International COSPAR satellite identification (yynnnaaa), consisting of the year (yy), launch number (nnn) and piece letter (aaa)	98067A	First catalog object of the 67th launch of the year 1998
	19–20	Epoch of the orbital element (year)	06	2006
	21–32	Epoch day and fraction of 24-hour day (UTC)	287.64456019	October 14, 15:28:10
	34–43	First time derivative of the mean anomaly (in [rev/d^2]) or ballistic coefficient *B*	.00008182	0.00008182
	45–52	Second time derivative of the mean anomaly (in [rev/d^3]) (Decimal between columns 45 and 46; exponent in columns 51–52)	00000-0	0.0
	54–61	Bstar/drag term *B** (in [1/$R_\oplus$]); (Decimal between columns 54 and 55; exponent in columns 60–61)	53355-4	$0.53355 \cdot 10^{-4}$
	63	Ephemeris type	0	SGP4 model
	65–68	Element number	860	860
	69	Check sum (modulo 10)	9	9
2	1	Line number identification (= 2)	2	Second line
	3–7	NORAD catalog number (example 16609)	25554	ISS (Zarya Module)
	9–16	*Inclination* (in [°])	051.6354	51.6354°
	18–25	Right ascension of ascending node (in [°])	292.0281	292.0281°
	27–33	Eccentricity with assumed leading decimal (between columns 26 and 27)	0013277	0.0013277
	35–42	Argument of the perigee (in [°])	096.0881	96.0881°
	44–51	Mean anomaly (in [°])	057.9543	057.9543°
	53–63	Mean motion (in [rev/d])	15.76874518	15.76874518°/d
	64–68	Revolution number at epoch	45200	45200
	69	Check sum (modulo 10)	8	8

Figure 2.2.12: *Breakdown of the NORAD two-line element set (example: the ISS).*

natural limit for orbital calculations produced with this model. Over the course of time one can recognize a quadratic increase in the error. This particularly affects the flight direction and can be traced back to an incorrect calculation of drag. Here the root is not so much the limits of the model but rather the poor predictability of atmospheric density. After a predictive period of a week the variance in the example above reaches 20 km, which represents a difference in the flyover time of almost 3 s. For higher satellite orbits

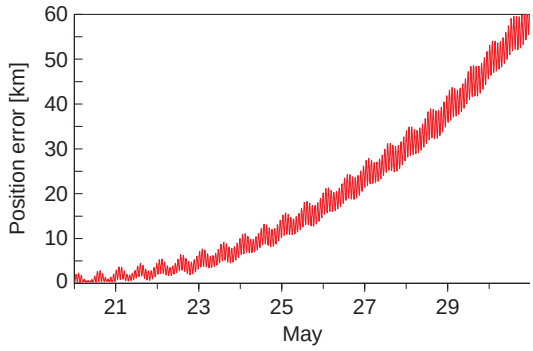


Figure 2.2.13: *Prediction accuracy of two-line elements for the satellite CHAMP in May 2001.*

(800 to 1300 km) the errors are typically significantly smaller. Two-line elements can be used effectively here even for predictions spanning several weeks.

2.2.1.5 Numerical Orbital Predictions

An alternative to the analytical models described above is the use of purely numerical methods to compute satellite orbits. Beginning with the known position and velocity at a specific time, numerical integration is used to predict the motion from one epoch to the next. A series of different single-step and multistep processes are available for this calculation, the basis of which will be introduced later in this section.

The choice of a suitable integration method depends upon the eccentricity of the orbit, the length of the orbital arc being handled, the required accuracy and several other factors. A "best" scheme is therefore impossible to name and users must always choose for themselves the most advantageous alternative. In general it is possible to note that higher order schemes are required to handle longer term predictions and to minimize the influence of rounding errors. The theoretical basis for these methods and their implementation in computer programs are extensively described in textbooks for numerical mathematics and related monographs (see [2.2.18]). For a comparative evaluation of numerical integration schemes for predicting satellite orbits, see [2.2.2] and [2.2.19].

Equation of Motion

The basis for the numerical integration is the description of the orbital motion by a suitable differential equation. Here we restrict ourselves to an equation of motion of the form $\dot{\boldsymbol{y}} = \boldsymbol{f}(t, \boldsymbol{y}(t))$ for the state vector

$$\boldsymbol{y}(t) = \begin{pmatrix} \boldsymbol{r}(t) \\ \boldsymbol{v}(t) \end{pmatrix} \tag{2.2.31}$$

composed of position and velocity.

The derivative

$$f(t, \boldsymbol{y}(t)) = \begin{pmatrix} \boldsymbol{v}(t) \\ \ddot{\boldsymbol{r}}(t, \boldsymbol{r}, \boldsymbol{v}) \end{pmatrix} \tag{2.2.32}$$

of the state vector in turn contains the velocity and acceleration vectors. In contrast to analytical orbit models requiring a closed-form analytic representation of the perturbation acceleration, numerical methods are subject to less stringent limitations. Here it is principally sufficient that the acceleration

$$\ddot{\boldsymbol{r}} = \boldsymbol{a}_{\mathrm{Grav}} + \boldsymbol{a}_{\mathrm{Sun}} + \boldsymbol{a}_{\mathrm{Moon}} + \boldsymbol{a}_{\mathrm{Drag}} + \boldsymbol{a}_{\mathrm{SolRad}} + \ldots \tag{2.2.33}$$

can be calculated pointwise and is progressively smooth enough. This is particularly important for modeling nongravitational forces whose calculation often involves interpolating table values.

Runge–Kutta Methods

The family of Runge–Kutta methods is among the best-known methods for the numerical integration of ordinary differential equations, and can be traced to the work of the mathematicians Carl Runge and Wilhelm Kutta. Starting with the **state vector** at time t, an approximate value at time $t + h$ is calculated as a **weighted average**

$$\boldsymbol{y}(t+h) \approx y(t) + h \sum_{i=1}^{m} b_i \boldsymbol{k}_i \tag{2.2.34}$$

of m derivatives

$$\boldsymbol{k}_i = \boldsymbol{f}\left(t + c_i h, \quad y(t) + h \sum_{j=1}^{i-1} a_{ij} \boldsymbol{k}_j\right) \tag{2.2.35}$$

at various midpoints. At each step one only requires the starting value $\boldsymbol{y}(t)$. At the same time the **step size** h may be freely chosen and adapted to current needs. The Runge–Kutta methods are therefore single-step methods.

The coefficients a_{ij}, b_i and c_i are chosen so that a best approximation of the true solution is produced.

In this context one speaks of an ***n*th-degree method** when the error of the numerical solution grows more slowly than h^{n+1} with the step size. In general this requires $m > n$ derivatives; only methods up to fourth degree are known for $m = n$. The coefficients are not uniquely determined by the order, meaning that there are indeed different Runge–Kutta methods of the same order.

For the most common, classical **Runge–Kutta fourth-order method**, the simple set of equations

$$\boldsymbol{y}(t+h) \approx \boldsymbol{y}(t) + \frac{h}{6}(\boldsymbol{k}_1 + 2\boldsymbol{k}_2 + 2\boldsymbol{k}_3 + \boldsymbol{k}_4) \qquad (2.2.36)$$

holds with

$$\boldsymbol{k}_1 = \boldsymbol{f}(t, \boldsymbol{y}(t))$$

$$\boldsymbol{k}_2 = \boldsymbol{f}(t + h/2, \boldsymbol{y}(t) + h\boldsymbol{k}_1/2)$$

$$\boldsymbol{k}_3 = \boldsymbol{f}(t + h/2, \boldsymbol{y}(t) + h\boldsymbol{k}_2/2)$$

$$\boldsymbol{k}_4 = \boldsymbol{f}(t + h, \boldsymbol{y}(t) + h\boldsymbol{k}_3)$$

For low-flying Earth satellites the application of this method lends itself to the use of step sizes between 30 and 60 s. A better approximation for each step can be achieved with substantially smaller values, but the accumulated rounding errors then dominate and soon lead to a worse overall result. For precise orbital predictions across many orbits the use of a higher order method (e.g., DOPRI8 from Dormand and Prince, see [2.2.18]) is indispensable. In contrast the RK4 method can be very advantageous for **real-time navigation** on-board a satellite when the predicted orbit is periodically updated with independent measurements [2.2.19].

Multistep Methods

Multistep methods follow an alternative approach to the integration of the equations of motion in which information from past epochs is included in the prediction of the next step. Besides the possibility to easily achieve higher orders, the necessary steps in multistep methods can be very simply organized. These methods were therefore used successfully very early in the calculation of planets and comets, and later assumed their natural place in the prediction of satellite orbits.

The basis of the multistep methods is the representation of the state vector as an integral with the form

$$\boldsymbol{y}(t+h) = \boldsymbol{y}(t) + \int_t^{t+h} \boldsymbol{f}(\tau, \boldsymbol{y}) \mathrm{d}\tau \qquad (2.2.37)$$

If the values of the state vector and therefore the derivative function are known for several previous times, the chronological sequence of $\boldsymbol{f}$ can be interpolated (and extrapolated) using a polynomial. The integral can then be easily evaluated and produces an approximate solution for the state vector $\boldsymbol{y}(t + h)$ as a function of $\boldsymbol{y}(t)$ and the known values $\boldsymbol{f}_i = \boldsymbol{f}(t + i\,h)$ with $i = 0, -1, -2, -3, \ldots$.

In the concrete case of the **Adams–Bashforth method** of the fourth order these equations are

$$\boldsymbol{y}(t+h) = \boldsymbol{y}(t) + \frac{h}{24} \cdot (-9\boldsymbol{f}_{i-3} + 37\boldsymbol{f}_{i-2} - 59\boldsymbol{f}_{i-1} + 55\boldsymbol{f}_i) \qquad (2.2.38)$$

They can be supplemented by a corrector of the fifth order stemming from Moulton:

$$\boldsymbol{y}(t+h) = \boldsymbol{y}(t) + \frac{h}{720}(-9\boldsymbol{f}_{i-3} + 106\boldsymbol{f}_{i-2} - 264\boldsymbol{f}_{i-1} + 646\boldsymbol{f}_i + 251\boldsymbol{f}_{i+1}) \qquad (2.2.39)$$

The resulting **predictor–corrector method** evaluates the derivative of the state vector twice at each location (and requires therefore twice the calculation effort per time step), but produces a much higher quality solution while enabling an estimation of the integration error at each step.

By reusing already known quantities, multistep methods are usually very effective. This is of particular advantage when – as in the case of a satellite orbit – the evaluation of the acceleration is very complex and requires costly calculations. However, it must be noted that the determination of the necessary starting values requires a complex starting calculation that complicates the use of these methods.

In the above description it was assumed that the integration is carried out with a constant step size. The coefficients of the Adams–Bashforth–Moulton method are then constants, making them particularly easy to use. As described in detail in [2.2.20], this is not a basic limitation. For example, the method developed by Shampine and Gordon supports differential

equation variable orders and step sizes and thus represents a powerful and flexible tool for dealing with many problems in orbital mechanics.

2.2.2 Orbit Determination

Given particular orbital elements or a starting state vector, the models covered in Section 2.2.1 make it possible to predict the orbit of a satellite. The task of orbit determination is to determine these parameters from observations of the satellite. Basic procedures for accomplishing this task were developed long before the first artificial satellite by various experts of celestial mechanics, with Karl Friedrich Gauss leading the way. In contrast to newly discovered planetoids or comets, the orbit of a satellite is usually not completely unknown. The nominal orbit is determined during mission planning before launch. Later the satellite is routinely monitored by mission control so approximate orbital information is always available. For this reason the orbit determination of satellites is usually concerned with improvements to approximated starting values and only in exceptional cases (e.g., military monitoring of intercontinental missiles or space debris) entails a classical initial orbit determination.

In the further course of this section the two most important mathematical methods for **improving orbital parameters** from satellite observations will be introduced: the **method of least squares** and **Kalman filtering**. Both methods are related at their core but pursue different goals and different applications have been developed for them. Parameter adjustment according to the method of least squares traditionally includes the collective processing of a large number of observations and is therefore primarily suited to postprocessing on the ground. The Kalman filter by contrast is particularly well suited to real-time processing, as only measurements performed at the current epoch are processed.

Before embarking upon a detailed description of these methods, it should first be explained which measuring instruments are available for observing and monitoring satellite orbits.

2.2.2.1 Tracking Systems

Today, predominantly **radiometric methods** are used for the orbital measurement of satellites. This involves using radio waves of various wavelengths to measure the direction to the satellite, its distance or its velocity along the line of sight [2.2.21]. Although optical cameras were used at the beginning of the space age to determine the position, today the use of telescopes is mostly limited to monitoring space debris and inactive satellites. Of great importance for high-precision orbit determination and geodetic research is the use of **laser ranging**, capable of measuring position with centimeter accuracy.

Ground Stations

Ground stations serve as the interface between the satellite and mission control (Figure 2.2.14). Besides relaying commands and receiving telemetry data, many of the deployed antennas can also be used for **tracking**.



Figure 2.2.14: *The 15 m S-band antenna (left) and 11 m Ku-band antenna (right) of the DLR ground station in Weilheim, Germany (Source: DLR).*

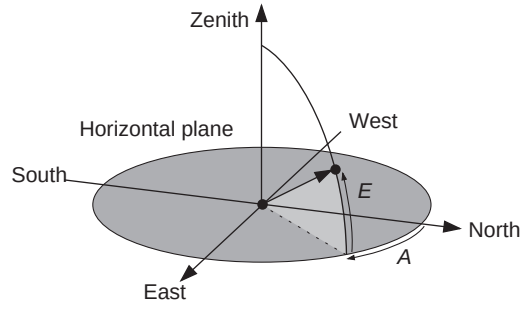


Figure *2.2.15: Definition of the direction angles azimuth (A) and elevation (E).*

In the simplest case, only the direction of the signal is determined. This is normally represented by the two angles of azimuth and elevation (Figure 2.2.15). The azimuth angle represents how far from north on the vertical axis the antenna must be turned to make contact with the satellite. The elevation describes the complementary angle above the horizon. Together these two values uniquely determine the direction to the satellite in a local east/north/zenith coordinate system.

To carry out **directional measurements**, nothing more than an active transmitter must be on-board the satellite. With the help of a large parabolic dish the transmitted signals are focused in the receiver antenna, also causing a pronounced directivity. Depending upon the antenna diameter D and the wavelength λ of the signal, the reception cone has a characteristic beam angle of

$$\varphi = 1.22 \frac{\lambda}{D} \tag{2.2.40}$$

around the line of sight of the antenna. Precise directional measurements therefore require a larger antenna and a short wavelength. The S-band, a part of the spectrum designated for communication with satellites, is composed of frequencies around 2 GHz and wavelengths of approx. 15 cm. The full width at half maximum (FWHM) of the reception cone for a 15 m antenna is then approximately 0.7°, which would only allow a very imprecise measurement. The **accuracy** can be significantly improved when several offset detectors are used in the antenna receiver instead of a single sensor. Instead of the (very flat) sensitivity maximum, the difference between the detector signals can be used to determine the direction. With the help of this method, known as monopulse radar (see also Section 6.3.4.5), the accuracy in the example above can be increased to better than 0.01°. This lies far below the nominal resolution of the antenna, but requires a good angular encoder and a very stable antenna mount.

At a distance of 600 km the angular measurement above is equivalent to a positional error of 100 m, and for a geostationary satellite this increases to 6 km. One remedy here is **distance measurement** (or ranging) implemented using the signal travel time. In contrast to angular measurements this is an active method that involves sending a signal from the ground that is returned by the satellite and then received back at the ground again. To make a travel time measurement possible, the carrier signal is modulated with a characteristic pattern of superimposed harmonic frequencies (tone ranging) or pseudorandom noise (pseudorandom noise ranging). This makes an accuracy from 1 to 10 m possible.

A so-called transponder is required on-board the satellite to receive the signal and return it delay free. To prevent interference or feedback, the signal is converted to a different frequency before it is returned. For example, in the S-band the frequencies from 2025 to 2120 MHz are designated for the ground to satellite (uplink) connection, while in the other direction (downlink) frequencies from 2200 to 2300 MHz are used. In accordance with international regulations the relation between the two frequencies is always 221/240.

As a third independent method, many ground stations can **measure the change in distance** based upon the **Doppler shift** of the received signal. If the distance between the sender and the receiver is increasing at a speed $\dot{\rho}$, a transmitted radio signal of frequency f will be received with a frequency shift of

$$\Delta f = -f \frac{\dot{\rho}}{c} \tag{2.2.41}$$

If the exact transmission frequency is known, it is possible to determine the velocity along the line of sight by measuring the Doppler shift. In the case of two-way measurements using a transponder, the current accuracy stands at around 0.1–1 mm/s. In principle, Doppler measurements are also possible without a signal broadcast from the ground (one-way measurement). In this case the measurement accuracy is strictly limited by the transmitter frequency stability on-board the satellite.

Satellite Laser Ranging

Distance measurement with the help of lasers (satellite laser ranging (SLR)) offers an extremely accurate alternative to microwave methods. It is used particularly for geophysical satellite missions where it serves not only as a means of measuring the orbit, but also for the precise determination of station movements. The operation of approx. 40 SLR stations worldwide is coordinated by the International Laser Ranging Service (ILRS) [2.2.22].

For SLR measurements a very short laser pulse is bundled with a telescope and sent to the satellite, see Figure 2.2.16. Here the light is returned by a corner



Figure 2.2.16: *A 1 m reflecting telescope for laser distance measurements at the Zimmerwald Observatory (Source: AIUB).*



Figure 2.2.17: *DORIS station on the island of Crozet (Source: C. Bricaud/IPEV).*

reflector with a "cat's eye" effect back to the ground and can be received there with the same telescope. The time between sending and receiving can be determined to within a few picoseconds. Using the known speed of light the distance can be determined. By aggregating the return time measurements from several pulses to averages, the SLR stations today can achieve accuracies to within the range of a few millimeters.

DORIS

The French system DORIS (Doppler Orbitography and Radiopositioning Integrated by Satellite) is based upon a worldwide network of **radio transmitters** (Figure 2.2.17) that transmits signals with high-frequency stability. Through comparisons to a reference oscillator on-board the satellite, the Doppler shift, and with it the orbit of the satellite, can be determined. As the measurements are first made on the satellite, DORIS also supports on-board determination of the orbit.

DORIS has been developed since the 1980s especially for the needs of altimeter missions that require highly accurate orbital measurements. Individual Doppler measurements with DORIS achieve an accuracy of about 0.5 mm/s. When performing a precise orbit determination, the altitude of a satellite above the Earth can be determined to within an accuracy of 1 cm. A **real-time navigation system** (DIODE) integrated in the DORIS receiver provides the position of the satellite with an accuracy of about 0.5 m [2.2.23] on-board.

GPS

GPS comprises a constellation of approximately 30 satellites that orbit the Earth at an altitude of 20 200 km with a period of 12 h [2.2.24]. The GPS system exists to provide users worldwide with an exact determination of their **position, velocity and time**. The constellation is conceived such that from every point on Earth (and near-Earth space) at least four GPS satellites can be received.

All the satellites use identical frequencies (L1: 1575.42 MHz; L2: 1227.6 MHz), but can be distinguished from one another by their characteristic "fingerprint." Using the signals it receives, a GPS receiver can calculate the difference between transmission and reception time. However, an error is introduced by inaccuracies in the timekeeping of the receiver. For this reason it is referred to as pseudo-distance measurement. Additionally each GPS satellite transmits **orbital elements** that allow the determination of its current position. If measurements from at least four satellites are available, the position and exact time can be deduced in the receiver. The current velocity of the receiver is also calculable from Doppler measurements.

In the mid-1980s GPS receivers were first used on low Earth orbit (LEO) satellites and have established themselves as an inexpensive and powerful **navigation system**. In contrast to the previously described tracking methods, GPS receivers not only provide individual position-dependent measurements, but make a direct determination of the position in space

possible at any given time. The availability of all measurements on-board the satellite represents an important building block of autonomous satellite navigation (Figure 2.2.18).

Comparable to an Earth-bound receiver, GPS receivers on LEO satellites today provide **instant position measurements** with an accuracy of about 10 m. In combination with a dynamic orbit determination using the elementary pseudo-distance and carrier phase measurements, much more exact orbital information can be achieved. For many Earth science missions today orbit determination accuracies of better than 10 cm are achieved. Using differential phase measurements for satellite formations, relative position accuracies in the millimeter range can be determined [2.2.25].

2.2.2.2 Observation Model

The goal of orbit determination (for improving the orbit) is to minimize the variation between a modeled orbit and observations. Section 2.2.1 described how the orbit of a satellite can be represented by a series of orbital elements or starting values, as well as by using various model parameters (e.g., radiation pressure coefficient or drag coefficient). Above and beyond this, for comparison to actual observations a **model of the measurement process** is necessary, unless direct three-dimensional position measurements are available (as is the case for a GPS receiver).

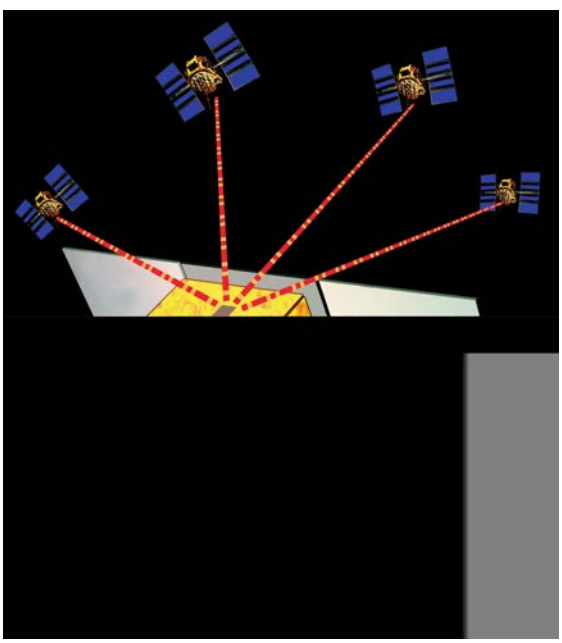


Figure 2.2.18: *Orbit determination of satellites with the help of the GPS system (Source: DLR).*

For the sake of the following discussion z represents a scalar (one-dimensional) measurement that depends on the orbit of the satellite. Accordingly the function $g(t, \boldsymbol{y}(t), \ldots)$ describes the measurement model and reflects the dependence of the measured quantity (e.g., the distance of the satellite) from the measurement epoch t, the state vector $\boldsymbol{y}(t)$ of the satellite at epoch t, as well as other conceivable parameters (e.g., systematic runtime delays). Alternatively, the measurement model can also be represented using the state vector $\boldsymbol{y}_0 = \boldsymbol{y}(t_0)$ for a starting epoch t_0 and force model parameter $\boldsymbol{p}$, which function, for purposes of discrimination, we refer to as $h(t, \boldsymbol{y}_0, \boldsymbol{p}, \ldots) = g(t, \boldsymbol{y}(t), \ldots)$. The residual $z - g$ or $z - h$ describes how well the modeled orbit fits the actual observations independently of the chosen representation.

Angle and Distance Measurements

As an example for deploying the measurement model the following discussion will examine the angle and distance measurements of a ground station. For this purpose a coordinate system will be used, the origin of which represents the antenna base, with the axes oriented to the east, north and zenith (see Figure 2.2.15). The position of the satellite in this so-called **topocentric coordinate system** can be generally represented in the form

$$\boldsymbol{s} = \boldsymbol{E} \cdot (\boldsymbol{U}(t) \cdot \boldsymbol{r}(t) - \boldsymbol{R}) \tag{2.2.42}$$

with $\boldsymbol{R}$ representing the coordinates of the ground station in an Earth-centered system (e.g., ICRF) and $\boldsymbol{U}$ the time-dependent transformation from the inertial to the Earth-fixed system. Furthermore,

$$\boldsymbol{E} = \begin{pmatrix} -\sin\lambda & +\cos\lambda & 0 \\ -\sin\varphi\cos\lambda & -\sin\varphi\cos\lambda & +\cos\varphi \\ +\cos\varphi\cos\lambda & +\cos\varphi\sin\lambda & +\sin\varphi \end{pmatrix} \tag{2.2.43}$$

describes the transformation in the local east, north and zenith system dependent upon the geographic latitude (λ) and longitude (φ) of the station.

If, on the other hand, one considers the distance ρ and the two angles of azimuth (A) and elevation (E), the relationship for the topocentric position vector holds as shown in Figure 2.2.15:

$$\boldsymbol{S} = \begin{pmatrix} S_O \\ S_N \\ S_Z \end{pmatrix} = \rho \cdot \begin{pmatrix} \cos E \ \sin A \\ \cos E \ \cos A \\ \sin E \end{pmatrix} \tag{2.2.44}$$

By solving for the individual measurements one then achieves the following **model function** for the distance and angle measurements:

$$\begin{aligned} g_\rho(t, r(t)) &= |\boldsymbol{S}|, \\ g_A(t, \boldsymbol{r}(t)) &= \arctan\,(s_O / S_N) \\ g_E(t, \boldsymbol{r}(t)) &= \arcsin\,(S_Z / |\boldsymbol{s}|) \end{aligned} \tag{2.2.45}$$

All measurements can then be represented by $\boldsymbol{s}(t, \boldsymbol{r}(t))$ as a function of the measurement epoch t and the related inertial satellite position $\boldsymbol{r}(t)$. Analogously, the functions can be formulated as $h_\rho\,(t, \boldsymbol{y}_0, \boldsymbol{p})$, $h_A(t, \boldsymbol{y}_0, \boldsymbol{p})$ and $h_E(t, \boldsymbol{y}_0, \boldsymbol{p})$, which describe the connection between the modeled observations and the starting state vector $\boldsymbol{y}_0$, as well as the force model parameters $\boldsymbol{p}$.

It should be stressed that various simplifications have been made in the representation above. To exactly model the measurement it would also be necessary to consider that the satellite moves during signal propagation. Its position at the time of reception at the ground station is therefore slightly different from its position at the time of transmission of the signal. Additionally, the signal experiences delays and deflections on its path through the atmosphere. These effects can be added to the measurement model through appropriate corrections without altering the basic concept.

2.2.2.3 Linearization

The model functions for the motion of the satellite and the description of individual measurements are generally too complex to allow directly solving for the $\boldsymbol{x} = (\boldsymbol{y}_0, \boldsymbol{p})$ unknown. It is therefore helpful to study only the effect of small offsets from a given **a priori value**.

In the case of actively operated satellites, an **approximate starting value** for the state vector $\boldsymbol{y}_0$ (or the orbital elements) for an epoch t_0 will be available. Such values are usually derived from predictions from earlier orbit determinations. Even at the start of a satellite mission, approximate orbital information is available based on the planned insertion orbit of the rocket. Likewise, meaningful starting values can also be found for the model parameters p (such as radiation pressure and drag coefficient) from earlier orbit determinations or theoretical considerations.

If these starting values are defined as $\boldsymbol{x}^{\text{ref}} = (\boldsymbol{y}_0^{\text{ref}}, \boldsymbol{p}^{\text{ref}})$, the model function for the measurement value h can then be developed in a Taylor series expanded around $\boldsymbol{x}^{\text{ref}}$ to the first order:

$$h(t, \boldsymbol{x}) = h(t, \boldsymbol{x}^{\text{ref}}) + \left(\frac{\partial h}{\partial \boldsymbol{x}}\right) \cdot (\boldsymbol{x} - \boldsymbol{x}^{\text{ref}}) + \cdots \tag{2.2.46}$$

The derivative

$$\frac{\partial h}{\partial \boldsymbol{x}} = \frac{\partial h\,(t, \boldsymbol{y}_0, \boldsymbol{p})}{\partial (\boldsymbol{y}_0, \boldsymbol{p})} = \frac{\partial \boldsymbol{g}\,(t, \boldsymbol{y}(t))}{\partial \boldsymbol{y}(t)} \cdot \begin{pmatrix} \dfrac{\partial \boldsymbol{y}(t)}{\partial \boldsymbol{y}_0} & \dfrac{\partial \boldsymbol{y}(t)}{\partial \boldsymbol{p}} \end{pmatrix} \tag{2.2.46a}$$

is thereby composed of three separate terms, namely:

- The derivative $\partial g / \partial \boldsymbol{y}$ of the modeled measurement with respect to the state vector at time t of the measurement
- The derivative $\boldsymbol{\Phi} = \partial \boldsymbol{y} / \partial \boldsymbol{y}_0$ of the state vector at the time of the measurement with respect to the state vector at the starting epoch t_0 (so-called field transfer matrix)
- The derivative $\boldsymbol{S} = \partial \boldsymbol{y} / \partial \boldsymbol{p}$ of the state vector at the time of the measurement with respect to the force model parameters (the so-called sensitivity matrix).

The **derivatives of the measurement model** can be obtained by the consecutive differentiation of the model function described above. At the same time, portions resulting from light transit time or other small corrections can usually be neglected. Derivatives of the satellite's velocity only occur with Doppler measurements, but not for distance or angle measurements. For example, for the distance measurement from a ground station one obtains the derivatives from the equations in the previous paragraph

$$\frac{\partial g_s}{\partial \boldsymbol{y}} = \begin{pmatrix} \dfrac{\partial g_s}{\partial \boldsymbol{r}} & \dfrac{\partial g_s}{\partial \boldsymbol{v}} \end{pmatrix} \tag{2.2.47}$$

with $\frac{\partial g_s}{\partial \boldsymbol{r}} = \frac{\partial g_s}{\partial \boldsymbol{s}} \cdot \frac{\partial \boldsymbol{s}}{\partial \boldsymbol{r}} = (s_O / s \; s_N / s \; s_z / s) \cdot E \cdot U$

as well as $\partial g_s / \partial \boldsymbol{v} = (0\ 0\ 0)$.

If the derivatives of the measurement model lead to unwieldy expressions, this is even more true for the calculations of the **state transition matrix** and the **sensitivity matrix**. If one wishes to correctly consider all the parts of the force model, both quantities must be generated by numerical integration of two matrix differential equations. For the equations of motion introduced in Section 2.2.1.5 these so-called **variational equations** are

$$\frac{\mathrm{d}}{\mathrm{d}t} \Phi(t) = \begin{pmatrix} 0 & 1 \\ \partial \boldsymbol{a} / \partial \boldsymbol{r} & \partial \boldsymbol{a} / \partial \boldsymbol{v} \end{pmatrix} \cdot \Phi(t) \qquad (2.2.48)$$

with $\Phi(t_0) = 1$ and

$$\frac{\mathrm{d}}{\mathrm{d}t} S(t) = \begin{pmatrix} 0 & 1 \\ \partial \boldsymbol{a} / \partial \boldsymbol{r} & \partial \boldsymbol{a} / \partial \boldsymbol{v} \end{pmatrix} \cdot S(t) + \begin{pmatrix} 0 \\ \partial \boldsymbol{a} / \partial \boldsymbol{p} \end{pmatrix} \qquad (2.2.49)$$

with $S(t_0) = 0$.

They in turn depend on the partial derivatives of the acceleration $\boldsymbol{a}$ of the position and velocity of the satellite, as well as the force model parameters. The exact calculation of these terms is explained at length in [2.2.2]. If one ignores the complexity of the resulting terms and the dimension of the problem, the numerical handling of the variation equations is performed according to the same principle and with the same methods as the integration of the equations of motion. To limit the calculation effort, when formulating the variation equations a series of simplifications are acceptable, for example a reduction in degree and order of the gravitational field.

2.2.2.4 Least Squares Adjustment

Estimation Problem

For parameter adjustment using the least squares method one attempts to find those model parameters for which the **sum of the squared differences** is minimized between the model and observations. Generally stated this is an optimization problem that can be reduced to an overdetermined system of equations through linearization around approximated starting values.

In its application to orbit determination problems this means that one begins by predicting the orbit of the satellite with the help of the starting values $\boldsymbol{x}^{\mathrm{ref}} = (\boldsymbol{y}_0^{\mathrm{ref}}, \boldsymbol{p}^{\mathrm{ref}})$ and for each measurement time calculates the associated modeled observation. From the differences between the measured and modeled observations (known as residuals) and the related partial derivatives, a correction $\Delta \boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^{\mathrm{ref}}$ can be calculated so that the sum

$$J = \sum_i (Z_i - h_i(t_i, \boldsymbol{x}))^2 = |\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})|^2 \qquad (2.2.50)$$

of the squared residuals for all observations reaches its smallest possible value (Figure 2.2.19). To simplify the further representation, all observations are collected in the vector $\boldsymbol{z} = (z_i)_{i=1,\ldots,n}$ of the measured values and a corresponding vector $\boldsymbol{h}$ of the modeled values.

If one only examines linear changes around the reference values then the task is to find the minimum of a quadratic function

$$J \approx |\Delta \boldsymbol{z} - \boldsymbol{H} \; \Delta \boldsymbol{x}|^2 = (\boldsymbol{H} \, \Delta \boldsymbol{x} - \Delta \boldsymbol{z})^T \; (\boldsymbol{H} \; \Delta \boldsymbol{x} - \Delta \boldsymbol{z}) \qquad (2.2.51)$$

of $\Delta \boldsymbol{x}$.

The **Jacobi matrix** $\boldsymbol{H} = \partial \boldsymbol{h} / \partial \boldsymbol{x}$ describes therein the derivative of the modeled observations with respect to the estimated parameters, while $\Delta \boldsymbol{z} = \boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x}^{\mathrm{ref}})$ represents the residue of the observations with regard to the reference orbit. From the minimum condition $\partial J / \partial \Delta \boldsymbol{x} = 0$ one finds the solution

$$\Delta \boldsymbol{x} = (\boldsymbol{H}^T \boldsymbol{H})^{-1} \; (\boldsymbol{H}^T \; \Delta \boldsymbol{z}) \qquad (2.2.52)$$

to the linearized problem.

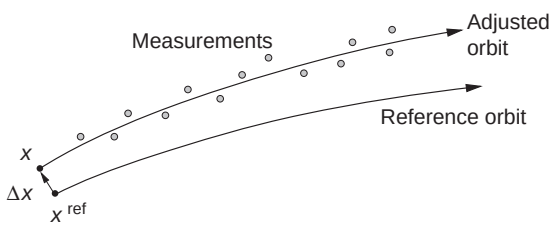


Figure 2.2.19: *Adjustment of the orbital parameters using the method of least squares.*

After applying this correction to the starting values, the modeled orbit will fit the observations significantly better. But using the simplifications that were made during linearization one rarely finds the desired minimum of the residual square sum in one step. The entire process can, however, be repeated several times to successively improve the result. Depending on the quality of the starting values, a handful of iterations is usually sufficient to determine the minimum of the residual square sum to better than 1% of its value.

It should be noted that the solution to the estimation problem can also be calculated without matrix inversion directly from the linear system of equations (the so-called **normal equations**):

$$(\boldsymbol{H}^T\ \boldsymbol{H})\ \Delta\boldsymbol{x}\ =\ (\boldsymbol{H}^T\ \Delta\boldsymbol{z}) \tag{2.2.53}$$

To achieve this, suitable numerical methods are available such as Gaussian elimination or LR transformation, which require significantly fewer calculations than a matrix inversion [2.2.26], [2.2.27]. This method is useful when only the solution to the equalization problem is required, but not when the covariance information described below is needed.

Weighting

In practice, it is often the case that in the course of orbit determination different types of observations (e.g., angle and distance measurements) or measurements of different quality (e.g., from two different stations) must be processed. This circumstance can be accounted for easily with a suitable weighting of the measurements. If one generally assumes that the i'th measurement exhibits a normally distributed error with a standard deviation of σ_i, the optimal solution to the orbital determination problem results from a minimization of the weighted residual square sum:

$$\begin{aligned} J\ &=\ \sum_i \left(\frac{z_i\ -\ h_i\,(t_i, \boldsymbol{x})}{\sigma_i} \right)^2 \\ &=\ (\boldsymbol{z}\ -\ \boldsymbol{h}(\boldsymbol{x}))^T\, \boldsymbol{W} (\boldsymbol{z}\ -\ \boldsymbol{h}(\boldsymbol{x})) \end{aligned} \tag{2.2.54}$$

The **weight matrix** $\boldsymbol{W} = \mathrm{diag}(1/\sigma_1^2, \ldots, 1/\sigma_n^2)$ is a diagonal matrix whose elements represent the inverse variance of the statistical measurement error. If weighting is considered when determining the minimum, one finds the following general relationship for the solution of the linear adjustment problem:

$$\Delta\boldsymbol{x}\ =\ (\boldsymbol{H}^T\ \boldsymbol{W}\ \boldsymbol{H})^{-1}\ (\boldsymbol{H}^T\ \boldsymbol{W}\ \Delta\boldsymbol{z}) \tag{2.2.55}$$

Covariance of the Solution

Besides the solution of the orbital determination problem itself, for many applications the expected **accuracy** is also of great interest. For this reason, in the preparation of a mission it must be decided which ground stations or measurement systems are necessary to determine the satellite orbit with a predetermined accuracy. These predictions can be derived from the **covariance matrix**. It is defined using the expected value

$$\boldsymbol{P}\ =\ \mathrm{E}\ ((\Delta\boldsymbol{x} -\ \overline{\Delta\boldsymbol{x}})\ (\Delta\boldsymbol{x}\ -\ \overline{\Delta\boldsymbol{x}})^T) \tag{2.2.56}$$

and primarily describes the scatter of the solution around its mean.

If one views measurement errors as random variables, then the resulting solution of the adjustment problem $\Delta\boldsymbol{x}$ itself is also a randomly distributed value. In the case of normally distributed measurement errors with a mean of zero, because of the linearity of the problem it first follows that the **expected value** $\overline{\Delta x} = \mathrm{E}\,(\Delta x)$ of the solution of the equalization problem in the mean always represents the true orbit. The orbit determination then exhibits no systematic error, as long as this is also true for the measurements themselves. It can further be demonstrated that the covariance of the solution is represented by the matrix

$$\boldsymbol{P}\ =\ (\boldsymbol{H}^T\ \boldsymbol{W}\ \boldsymbol{H})^{-1} \tag{2.2.57}$$

that was already determined to be a part of the least squares adjustment problem.

As a simple measure of the accuracy of the orbit determination one can initially examine the **diagonal elements** $P_{ii} = \sigma^2(x_i)$ of the covariance matrix. These contain the squares of the **standard deviation** of the individual estimation parameters. They represent, for example, the uncertainty of the starting position vector or the estimated semi-major axis. These error calculations only include the assumed accuracy of the measurements (through the weight matrix $\boldsymbol{W}$) and the type of measurement and their distribution (through the derivative matrix $\boldsymbol{H}$). The covariance matrix can therefore only provide statistical results but does not cover the influence of individual stray values or systematic measurement errors. Correct

interpretation of the results therefore requires appropriate caution and experience. As a rule it is possible to recognize if the orbit is well determined by the observations or if individual parameters cannot be adequately determined due to poor distribution of the observations.

A priori Information

In general the number of measurements is significantly greater than the number of unknowns. Principally one has an **overdetermined system of equations** whose solution is determined by the method of least squares. Unfortunately this overdetermination does not always guarantee that the values of all estimation parameters were reliably estimated. For example, if one only has measurements from a single satellite flyover of the ground station, the semi-major axis can only be fixed insufficiently. If, by contrast, measurements from two successive passes are available, then the orbital period, and thereby the major semiaxis, are very well determined. The quality of the orbit determination is therefore less dependent on simply the number of measurements than on a good distribution across time and location. Where this cannot be achieved, it is possible to incorporate independent information in the parameter estimation. For example, if the variance with which a satellite is injected into orbit by its rocket is known, this information can be considered along with the actual measurements when determining the orbit.

This additional information is generally described completely by the **a priori value** $\boldsymbol{x}^{\mathrm{ap}} = (\boldsymbol{y}_0^{\mathrm{ap}}, \boldsymbol{p}^{\mathrm{ap}})$, of the estimation parameters, and the associated covariance $\boldsymbol{P}^{\mathrm{ap}}$. By minimizing the combined cost function

$$J = |\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})|^2 + (\boldsymbol{x} - \boldsymbol{x}^{\mathrm{ap}})^T \boldsymbol{P}^{\mathrm{ap}} (\boldsymbol{x} - \boldsymbol{x}^{\mathrm{ap}}) \quad (2.2.58)$$

one obtains the solution

$$\Delta\boldsymbol{x} = (\boldsymbol{H}^T \boldsymbol{W} \boldsymbol{H} + \boldsymbol{P}^{\mathrm{ap}})^{-1} \cdot (\boldsymbol{H}^T \boldsymbol{W} \Delta\boldsymbol{z} + \boldsymbol{P}^{\mathrm{ap}} (\Delta\boldsymbol{x} - \Delta\boldsymbol{x}^{\mathrm{ap}})) \quad (2.2.59)$$

to the orbit determination problem with a priori information. Similar to the simple adjustment problem, the covariance of this solution is again described by the normal equation matrix

$$\boldsymbol{P} = (\boldsymbol{H}^T \boldsymbol{W} \boldsymbol{H} + \boldsymbol{P}^{\mathrm{ap}})^{-1} \quad (2.2.60)$$

A priori information and measurements are treated equally in this formulation. This also makes it possible to combine several consecutive data sets without processing all measurements simultaneously. Here the result of a previous orbit determination is entered as a priori information in the following orbit determination. The a priori values and their covariance then describe the information about the orbit contained in all previous measurements. At the end, an orbit determination results that takes into account information from a large data set, even if only a small number of new measurements were directly processed.

2.2.2.5 Kalman Filter

If one further elaborates the thoughts above, an alternate form of orbit determination results, known as the Kalman filter. There, observations are no longer processed in large batches but rather **epoch by epoch**. At the same time, estimation parameters (like orbital elements or state vector) do not refer to a chosen starting epoch but in each case to the time of the last measurement. The Kalman filter is therefore particularly of interest for **real-time processing** and is primarily used in on-board navigation systems.

The Kalman filter process is composed of two elementary steps:

- Starting from the last known value of the state vector and its covariance, both values are first predicted to the time of the next measurement in the **time update**.
- Then in the **measurement update**, an optimal estimate of the state vector and its covariance for the new epoch is determined based upon this a priori information and current observations.

Both steps are performed successively and repeated continuously (Figure 2.2.20).

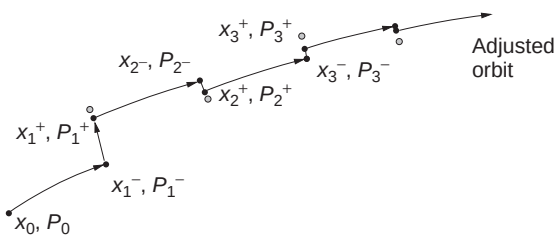


Figure 2.2.20: *Orbit determination using a Kalman filter.*

The starting point for the time update is the information of the state $\boldsymbol{x}_i^+$ and the corresponding covariance $\boldsymbol{p}_i^+$ at epoch t_i. The notation "+" indicates that it refers to the value *after* a previous measurement update. By contrast, values *before* the measurement update are indicated by a "–" sign With the help of the known force models and a numerical integration method the state vector can then be propagated to the next epoch t_{i+1}. In addition to the state vector $\boldsymbol{x}_{i+1}^-$ the integration of the variational equation additionally provides the **state transition matrix**

$$\boldsymbol{\Phi}_{i+1} = \partial \boldsymbol{x}_{i+1} / \partial \boldsymbol{x}_i \tag{2.2.61}$$

This indicates how small changes to the state vector at the start of the prediction interval affect the state vector at the end of the interval, and serves to linearly propagate the scatter of the starting values. In this way one obtains the predicted covariance

$$\boldsymbol{P}_{i+1}^- = \boldsymbol{\Phi}_{i+1}\, \boldsymbol{P}_i^- \boldsymbol{\Phi}_{i+1}^T \tag{2.2.62}$$

that describes the expected uncertainty of the state vector before the processing of new measurements. As a special feature, at this point the Kalman filter offers the possibility of also considering **statistical errors** in the prediction itself. For this, the covariance is increased by an amount $\boldsymbol{Q}$ that describes the so-called **process noise** and takes account of errors in the force model or the numerical integration:

$$\boldsymbol{P}_{i+1}^- = \boldsymbol{\Phi}_{i+1}\, \boldsymbol{P}_i^- \boldsymbol{\Phi}_{i+1}^T + \boldsymbol{Q}_i \tag{2.2.63}$$

Small values of $\boldsymbol{Q}$ reflect a high-quality orbit model and allow the filter to smooth the orbit across many epochs. By contrast, a high degree of noise means that the estimated orbit is always determined by the newest measurements and evidences a large statistical uncertainty. The "memory" of the filter can therefore be controlled by the choice of process noise and thus be adapted to the obtaining circumstances. Usually, however, a series of simulations and instinct are required to optimally use a Kalman filter.

For the most part, the measurement update of the Kalman filter is equivalent to the estimation with a priori information described in the last section. In the course of this procedure, the predicted state vector and its covariance are improved by the addition of new measurements:

$$x_{i+1}^+ = \boldsymbol{x}_{i+1}^- + \boldsymbol{K}_{i+1} \cdot (z_{i+1} - g_{i+1}) \tag{2.2.64}$$

How strongly the state vector is thereby corrected depends for one thing on the residual $z_{i+1} - g_{i+1}$ of the observation, and for another on the relative weighting of a priori information and observation. This weighting determines the so-called **Kalman gain** of the filter that results in

$$\boldsymbol{K}_{i+1} = \boldsymbol{P}_{i+1}^- \boldsymbol{G}_{i+1} \cdot (\sigma_{i+1}^2 + \boldsymbol{G}_{i+1}\, \boldsymbol{P}_{i+1}^- \boldsymbol{G}_{i+1}^T)^{-1} \tag{2.2.65}$$

The row vector $\boldsymbol{G}_{i+1} = \partial g_{i+1} / \partial x_{i+1}$ describes the derivative of the modeled observations according to the current state vector, and σ_{i+1} the standard deviation of the measurement error. Additionally the covariance can be expressed as

$$\boldsymbol{P}_{i+1}^+ = (1 - \boldsymbol{K}_{i+1}\, \boldsymbol{G}_{i+1}) \cdot \boldsymbol{P}_{i+1}^- \tag{2.2.66}$$

with the help of the Kalman gain. Although the formulation of the measurement update in the Kalman filter bears little resemblance to the estimation with a priori information introduced in Section 2.2.2.4, these two expressions are mathematically identical. However, the Kalman filter does not require any large matrix operations for updating the estimation.

Characteristic of the Kalman filter is not only the epoch-by-epoch processing of individual measurements, but also the estimation of the state vector only for the current time. The Kalman filter thus distinguishes itself from classical orbit determination using the least squares method in which the orbital elements or the position and velocity are determined for a chosen starting epoch. One disadvantage of sequential processing is the circumstance that only past or current, but naturally no future, measurements are included in the estimation of the current state vector. By contrast, the method of least squares provides an optimal fitting of all observations of a given data set. For this reason it is generally preferred for the *post facto* orbit determination on the ground where no serious real-time demands exist.

2.2.3 Orbit Design and Station Keeping

The third and last section of this discussion of orbital mechanics deals with designing and maintaining the

orbit of a satellite in Earth orbit. After a discussion of simple maneuvers for the transfer between different orbits, the especially important practical cases of Sun-synchronous remote sensing satellites and geostationary communications satellites will be examined more closely.

2.2.3.1 Hohmann Transfer

For many applications, achieving and maintaining a predetermined orbit is a fundamental prerequisite for mission success. One example is communications satellites which can only carry out their desired geosynchronous motion at a specific altitude above the equator. Remote sensing satellites must also hold predetermined orbital inclinations and orbit altitudes to maintain the required lighting conditions and ground tracks.

Due to the injection dispersion of current launch vehicles and inevitable orbital perturbations, a predetermined ideal orbit cannot be reached and maintained right away. For this reason so-called **orbital maneuvers** are used to correct variations from the nominal orbit. These are performed utilizing a system of thrusters to modify the velocity of the satellite $\boldsymbol{v}$ by a desired value $\Delta\boldsymbol{v}$. In the ideal case these changes are made so quickly that the length of the maneuver is negligible compared to the orbital period. After such an impulse-type maneuver the satellite flies in a new orbit characterized by new orbital elements. The old and new orbits share the point $\boldsymbol{r}$ at which the maneuver was carried out. As a result of the maneuver only three orbital elements (or three linearly independent combinations of orbital elements) are changed, while the remaining orbital parameters remain constant. If a correction of all six orbital elements is needed, at least two independent, time-offset maneuvers are required.

It is possible to discriminate between two types of maneuvers. If the change in velocity $\Delta\boldsymbol{v}$ lies in the orbital plane, the size and form of the orbit (a, e) and perigee (ω, M) are affected. Alternatively, inclination and the line of nodes (i, Ω) are altered by impulses perpendicular to the orbital plane.

Important special cases are **thrust and braking maneuvers** executed directly at perigee or apogee. Independent of the eccentricity of a satellite's orbit, the motion at the nearest and furthest point to the Earth is always perpendicular to the radius of the orbit. A change in the velocity vector in the direction of flight (thrust) or against the direction of flight (braking) only affects the semi-major axis and the eccentricity of the orbit.

In detail, for the velocities at perigee (v_p) and apogee (v_a) the relationships

$$v_p = \sqrt{GM_\oplus \left(\frac{2}{r_p} - \frac{2}{r_p + r_a} \right)} \tag{2.2.67}$$

and

$$v_a = \sqrt{GM_\oplus \left(\frac{2}{r_a} - \frac{2}{r_p + r_a} \right)} \tag{2.2.68}$$

hold, based upon which simple orbital maneuvers can be calculated.

A well-known example is **Hohmann transfer** for moving between two circular orbits. For example, to raise a satellite from LEO to a geostationary orbit (Figure 2.2.21) its velocity must first be increased from 7.7 km/s to about 10 km/s. This additional energy allows the satellite to reach an altitude of approximately 36 000 km along an elliptical path. At

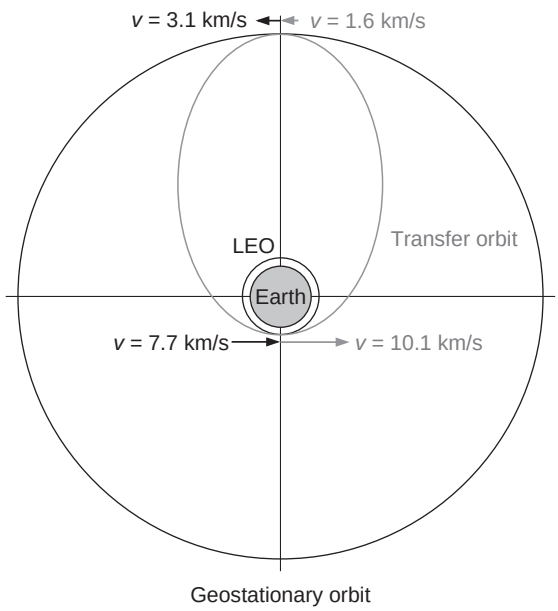


Figure 2.2.21: *Hohmann transfer to a geostationary orbit.*

apogee its velocity then falls to 1.6 km/s. If the velocity is again increased by 1.4 km/s, the orbit is raised at the opposite point and thereby achieves the desired circular form.

In practice, geostationary satellites are normally injected directly into the highly elliptical transfer orbit so that only apogee maneuvers are necessary. These are usually divided into several steps to keep individual maneuvers short.

In the example above it might at first be mystifying that two increases in velocity caused by the maneuvers result in a lower orbital velocity. This paradox is explained by the increase in potential energy achieved by increasing the orbital altitude, but in no way makes an intuitive prediction of the effects of thrust and braking maneuvers easier.

2.2.3.2 Remote Sensing Satellites

A large number of satellites in LEO regularly monitor and map the Earth's surface across a diverse spectral range. Here it is normal to synchronize the orbital period with the rotation of the Earth so that after a certain period of time the satellite flies over the same ground track. This is known as a "repeat orbit." For optical observations it is also important to ensure that lighting conditions vary as little as possible between different exposures of the same area. This is known as a "Sun-synchronous orbit." These factors present the conditions that must be taken into account during mission analysis when choosing the orbit's semi-major axis and inclination.

If special measures were not taken, completely different orientations of the Sun to the satellite orbit would result over the course of a year. This is due both to the annual orbit of the Earth around the Sun and to the node drift caused by oblateness, the rotational flattening of the Earth. In one year the right ascension of the Sun changes by 360°. For an inertially fixed orbit, the angle between the line of nodes and the direction of the Sun projected onto the equator grows daily by about 1° (or precisely 0.985 647°). When planning the orbit of a remote sensing satellite one can use the circumstance that the orbital plane itself exhibits a prograde (meaning forward-facing) rotation when the orbital inclination is greater than 90°. With an inclination of approx. 98°, the daily node drift for orbits between 700 and 900 km altitude is just high enough that they balance out the change in the direction of the Sun. Such Sun-synchronous orbits make possible (nearly) constant lighting conditions when flying over areas of the same geographic latitude. In particular, Sun-synchronous orbits distinguish themselves by always passing over the equator at the same local time. As an example, a local time of 10:30 is common. This ensures a sufficiently high position of the Sun (and with it enough light for the photographic exposure) while producing enough shadows to guarantee contrast-rich photographs. By nature, Sun-synchronous satellites travel along nearly polar orbits. They also enjoy nearly optimal coverage of a large portion of the Earth. Table 2.2.2 contains a collection of examples of remote sensing satellites in Sun-synchronous orbits. Besides information about orbital altitude and period, it also contains the satellite's ground track repeat period and the local time of ascending node crossing.

From the relationships for **secular node drift** introduced in Section 2.2.1.3 for small eccentricities ($e \gg 1$), the condition

$$\Delta\dot{\Omega} = -\frac{3}{2} J_2 \, n \, \frac{R_\oplus^2}{a^2} \cos(i) = 0.985647°/\mathrm{d} \qquad (2.2.69)$$

for a Sun-synchronous orbit results. For a given semi-major axis it is then possible to determine the

Table 2.2.2: *Orbital parameters for various Sun-synchronous remote sensing satellites.*

	ALOS	Landsat 5	ENVISAT	SPOT	IRS-1A
Altitude (a)	692 km	705 km	800 km	832 km	904 km
Orbital period (T)	98.5 min	99 min	100.6 min	101 min	103.2 min
Inclination (i)	98.16°	98.2°	98.55°	98.7°	99°
Repeat period (K days)	46	16	35	2	22
Orbits per repeat period (N)	671	233	501	369	307
Local time of the node transit	10:30	09:30	10:00	10:30	10:00

orbital inclination for which the node drift exactly compensates for the daily movement of the Sun along the equator.

The orbital altitude or semi-major axis itself is fixed by the desired orbital period or the required spacing of neighboring ground tracks for a **repeat orbit**. If the **draconitic period** (the time between consecutive equator crossings) is denoted as T_Ω, then the distance between neighboring ground tracks in east longitude is given by

$$\Delta\lambda = (\Delta\dot{\Omega} - \dot{\Theta}) \cdot T_\Omega \tag{2.2.70}$$

To ensure that after N orbits a repeating pattern of ground tracks results which cover the Earth K times, the condition

$$N \cdot (\Delta\dot{\Omega} - \dot{\Theta}) \cdot T_\Omega = -K \cdot 360° \tag{2.2.71}$$

for the orbital period T_Ω must be met. However, it must be considered that the mean draconitic motion

$$\eta_\Omega = n + \Delta\dot{\omega} + \Delta\dot{M} \tag{2.2.72}$$

differs due to secular changes in perigee and the mean anomaly by the value

$$\Delta\dot{\omega} + \Delta\dot{M} \approx +\frac{3}{2} J_2 \, n \frac{R_\oplus^2}{a^2} \left[4\cos^2(i) - 1 \right] \tag{2.2.73}$$

from the unperturbed mean motion

$$n = \sqrt{\frac{GM_\oplus}{a^3}} \tag{2.2.74}$$

Considering this difference, using the equations above it is also possible to derive the Keplerian orbital period and thereby the desired semi-major axis a.

If after orbital injection the altitude and eccentricity do not exactly conform to specifications, both values can subsequently be corrected by maneuvers in or against the flight direction. The method is fundamentally the same as previously described for the Hohmann transfer.

Similarly, the orbital inclination and the node position can be adjusted to match the mission requirements using maneuvers perpendicular to the direction of flight, see Figure 2.2.22. An **impulse-type maneuver** of magnitude Δv_N in the direction of the orbit normal causes a rotation of the orbit plane by an angle

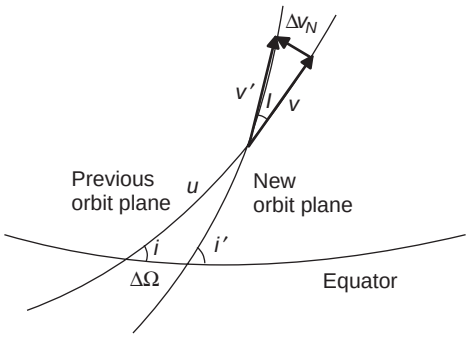


Figure 2.2.22: *Inclination maneuvers.*

$$I = \arctan(\Delta v_N / v) \approx \Delta v_N / v \tag{2.2.75}$$

For an orbital velocity of approx. 7500 m/s a maneuver of about 130 m/s would be required to change the orbital plane by 1°.

The impulse perpendicular to the orbital plane affects the orbital inclination or the node position depending upon the argument of latitude at which the maneuver is carried out. To a first approximation, the following applies for a circular orbit:

$$\begin{aligned} \Delta i &= \frac{\Delta v_N}{v} \cdot \cos(u) \\ \Delta\Omega \cdot \sin i &= \frac{\Delta v_N}{v} \cdot \sin(u) \end{aligned} \tag{2.2.76}$$

Therefore, an exclusive **change in inclination** results only from maneuvers made directly in the ascending (or descending) node.

Even after reaching the nominal mission orbit, smaller **correction maneuvers** are required fairly regularly to compensate for natural orbital perturbations. This is first and foremost related to the effects of **drag**, which leads to a slow decrease in altitude and thereby to a change in the orbital period. As a result, the ground track of remote sensing satellites is shifted, which can usually only be tolerated within a margin of about 1 km.

Under the simplified assumption of **constant drag** the major semi-major axis decreases

$$a(t) = a_0 + \dot{a}_0 \, (t - t_0) \tag{2.2.77}$$

with time at the constant rate of

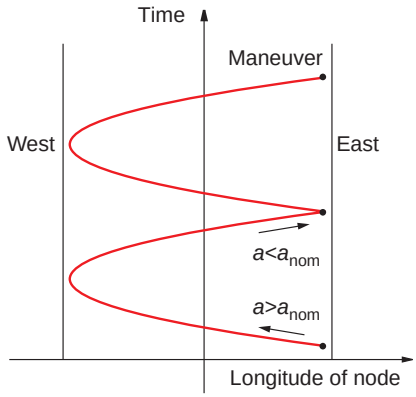


Figure 2.2.23: *Control of the ascending node of a remote sensing satellite.*

$$\dot{a}_0 = -C_D \frac{A}{m} \rho \cdot n\, a^2 \qquad (2.2.78)$$

The linear decrease in the semiaxis then leads to a continuous reduction of the time between consecutive equatorial crossings and thus to a quadratic shift in the ascending node to the east (Figure 2.2.23). If the ground track of a satellite has shifted too far from the nominal position, this drift can be stopped and reversed by a maneuver to raise the orbit. Ideally this maneuver is planned so that the apex of the node drift lies at the western window boundary, thus maximizing the time between maneuvers.

In practice the assumption of constant atmospheric drag is not sufficient. The density of the atmosphere can undergo significant variations depending on the solar activity. The actual motion of the ascending satellite thus diverges from the idealized parabolic form and must be corrected with additional or alternative maneuvers. This is illustrated using the example of the TOPEX satellite in Figure 2.2.24.

The interest in **on-board orbital control** has grown with the availability of GPS-based navigational data on many remote sensing satellites. Besides simplifying ground control, in combination with a properly designed propulsion system it can also improve control accuracy. Successful experiments with automatic orbital control for the EO-1 and Demeter satellites are described in [2.2.30] and [2.2.31].

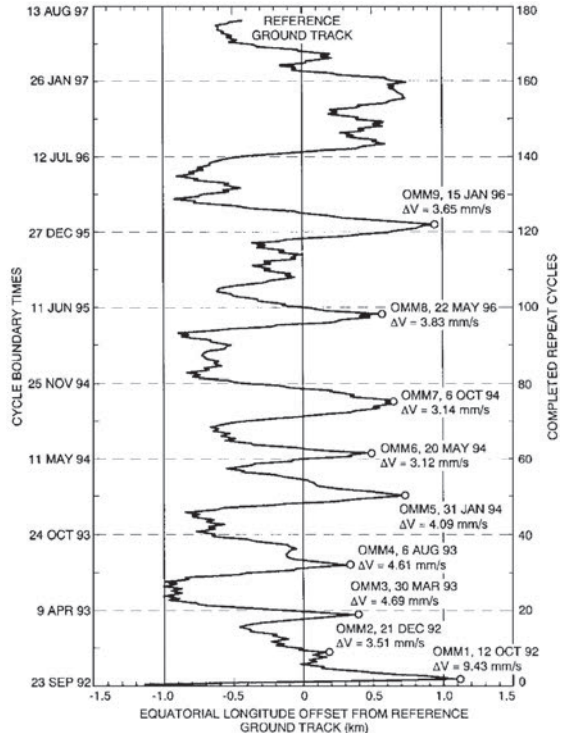


Figure 2.2.24: *Control of the ground track of the TOPEX satellite (Source: from [2.2.29], © AIAA).*

2.2.3.3 Geostationary Satellites

Geostationary satellites are characterized by the fact that they almost exactly maintain their position over a given point on the Earth's surface. The prerequisite is that they possess a nearly circular equatorial orbit and that their orbital period is the same as the length of one Earth rotation. From Kepler's third law, this results in a semi-major axis of

$$a = \sqrt[3]{GM_\oplus \left(\frac{23^{\mathrm{h}}\, 56^{\mathrm{m}}}{2\,\pi} \right)^2} = 42\ 162.8\ \mathrm{km} \qquad (2.2.79)$$

The radial attraction at the equator is slightly weaker than for a perfectly spherical Earth due to oblation. In fact, the **geosynchronous orbital radius** is approximately 1.5 km larger, being $a_{geo} = 42\ 164.3$ km.

For the description of a (nearly) geostationary satellite orbit it is helpful to use six alternative parameters in place of the classical Keplerian orbital elements.

First there is the **mean geographic longitude** $l = (\Omega + \omega + M) - \Theta$ of the satellite, composed of the right ascension of the ascending node (Ω), the argument of latitude (ω) and the mean anomaly (M), as well as the sidereal time. This is supplemented by the **drift rate**, defined as the time derivative of the mean longitude that replaces the major semiaxis. Deviations from an ideal circular orbit are described by the **eccentricity vector**:

$$\boldsymbol{e} = \begin{pmatrix} e_x \\ e_y \end{pmatrix} = e \cdot \begin{pmatrix} \cos(\Omega + \omega) \\ \sin(\Omega + \omega) \end{pmatrix} \tag{2.2.80}$$

Its magnitude matches the eccentricity of the orbit, and its direction expresses the right ascension of the perigee. The **inclination vector** is similarly formed as

$$\boldsymbol{i} = \begin{pmatrix} i_x \\ i_y \end{pmatrix} = i \cdot \begin{pmatrix} \cos\Omega \\ \sin\Omega \end{pmatrix} \tag{2.2.81}$$

It is a measure of the orbital inclination and the direction of the line of nodes.

In the ideal case of a strictly geostationary satellite, l would be constant and equal to the geographical latitude assigned to the satellite. The other elements would be zero for a geosynchronous circular orbit in the plane of the equator.

In reality this can never be achieved since various periodic and long-term **perturbations** continually change the orbit [2.2.32], [2.2.33]:

- The **irregular form of the Earth** causes a slow acceleration of a geostationary satellite along its orbit. The magnitude and sign of this acceleration depend on the satellite's location above the equator. For example, maximum values of $\pm 0.002°/d^2$ are reached over South America, East Africa and Indonesia. At an east longitude of 75°, −105°, 162° and −12° the acceleration disappears (Figure 2.2.25), but only the first two points are stable.
- Due to **gravitational perturbations** from the Earth, Sun and Moon, the eccentricity vector exhibits periodic variations around its mean. These have a typical amplitude of 10^{-4} on a time scale of one day and one month. The mean eccentricity is also subject to the long-term influence of radiation pressure, leading to a circular motion of the eccentricity vector in the (e_x, e_y) plane. The radius of this circle depends on the relationship between area and mass, and for modern satellites with large solar panels can reach values up to $7 \cdot 10^{-4}$.
- The **inclination vector** exhibits a **mean drift** of 0.75–0.95° per year in the direction $+i_y$ with a superimposed periodic variation of two weeks and six months in both axes. These perturbations are caused by the attraction of the Sun and Moon which impart a constant torque on the satellite orbit and so change the orbit normal.

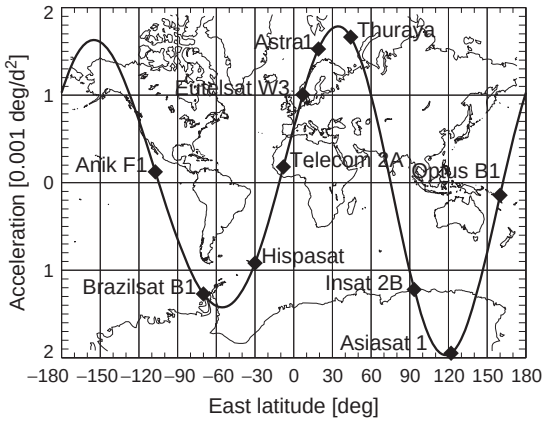


Figure 2.2.25: *Acceleration of a geostationary satellite in the direction of east longitude (Source: DLR).*

In accordance with international agreements, most geostationary satellites must be kept in a narrow window whose magnitude in geographical longitude and latitude may not exceed ±0.1°. This prevents mutual disturbances of transmission paths and allows controlled usage of the limited space in the geostationary ring.

If steps were not taken, various perturbations would cause the geostationary satellite to leave its control box of approximately 150 km × 150 km within a few weeks. To prevent this, **orbital control maneuvers** are carried out in a regular cycle of about two weeks. North–south maneuvers serve to control the inclination vector (the orbital plane), and east–west maneuvers steer the eccentricity and drift of the satellite.

The effect of impulsive maneuvers in radial (R), easterly (E) and northerly (N) directions on the orbital elements of a geostationary satellite are described to a first approximation by the equations [2.2.34].

2

$$
\begin{aligned}
\Delta D &= -\frac{3}{a}\,\Delta v_{\mathrm{E}} \\
\Delta l &= -\frac{2}{v}\,\Delta v_{\mathrm{R}} + \tau \cdot \Delta D \\
\Delta e_x &= \frac{1}{v}\,(+\Delta v_{\mathrm{R}} \sin\alpha + 2\,\Delta v_{\mathrm{E}} \cos\alpha) \\
\Delta e_y &= \frac{1}{v}\,(-\Delta v_{\mathrm{R}} \cos\alpha + 2\,\Delta v_{\mathrm{E}} \sin\alpha) \\
\Delta i_x &= \frac{1}{v}\,\Delta v_{\mathrm{N}} \cos\alpha \\
\Delta i_y &= \frac{1}{v}\,\Delta v_{\mathrm{N}} \sin\alpha
\end{aligned}
\qquad (2.2.82)
$$

Here, a = 42 164 km (the orbital radius) and v = 3.0 km/s (the satellite's velocity). The angle α stands for the right ascension of the satellite during the maneuver and τ for the time passed since the maneuver.

It is easy to recognize that thrust in a north–south direction only affects the inclination vector. The remaining elements can be corrected both with radial and with east–west maneuvers. As radial changes in velocity are less effective, thrusters in this direction are normally not used. However, the placement of antennas or solar panels often forces a suboptimal thruster orientation. In such cases the actual thrust vectors must be included when planning maneuvers.

An example is illustrated in Figure 2.2.26 for the **maneuver planning** results of a fictitious satellite at longitude 110°E. With an assumed mass of 700 kg and a cross-sectional area of 10 m^2, the radius of the

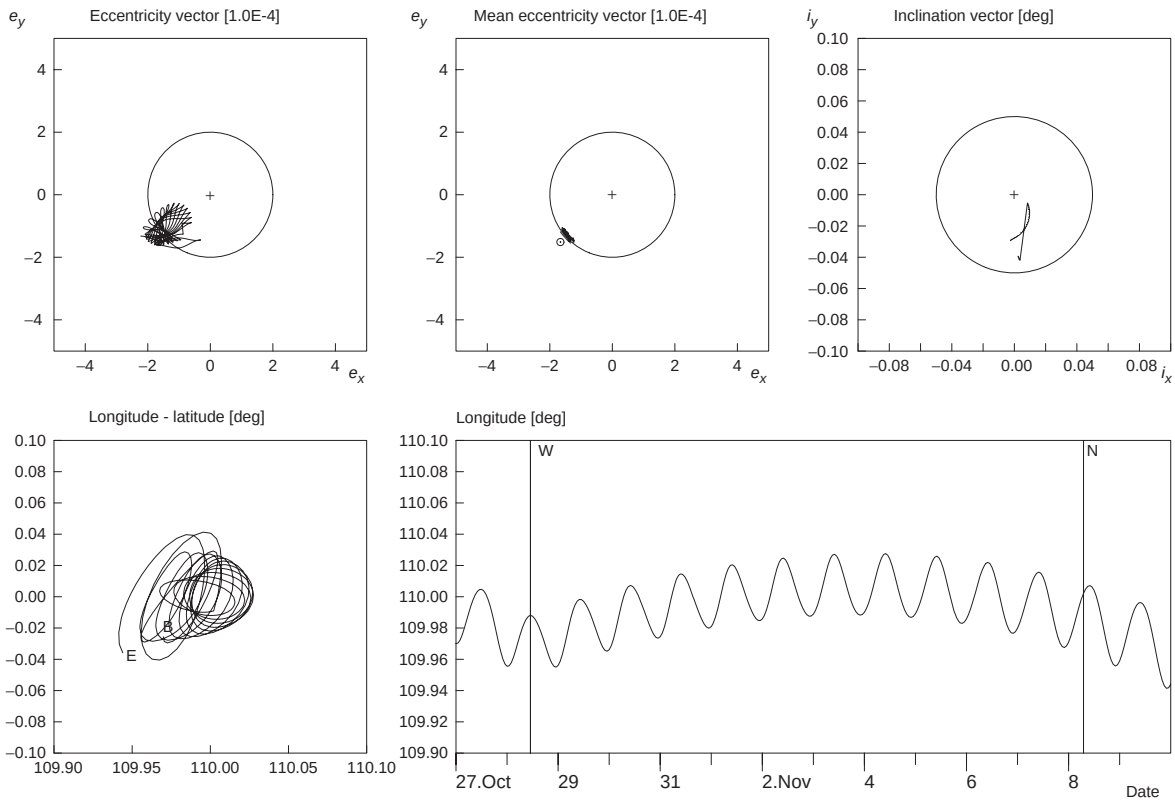


Figure 2.2.26: *Orbital control of a geostationary satellite at longitude 110°E. The top of the figure shows the change in the osculating and mean eccentricity vectors (left and middle), as well as the inclination vector (right). In the bottom section, the movements of the satellite in the control window and the variation over time of the geographic longitude are illustrated. Also labeled are the time of a west maneuver for longitudinal control and a north maneuver to control the orbital inclination (Source: DLR).*

natural eccentricity circle is only $2 \cdot 10^{-4}$ rad. This entails a small daily periodic oscillation of the satellite with an amplitude of 0.011° in an east–west direction. To control the longitude, in this case a simple west maneuver of 8 cm/s is sufficient. This slightly reduces the semi-major axis and causes a drift to the east, which is later canceled out by the natural perturbation of the Earth's gravitational field. Near the end of the 14-day station-keeping cycle, an additional north maneuver of 2 m/s is carried out to compensate for the drift of the inclination vector under the influence of the Sun and Moon. The inclination of the satellite orbit relative to the plane of the equator and therefore the daily latitude variation can thus be held to under 0.05°.

For a more detailed account of maneuver planning for geostationary satellites referral to the appropriate textbooks ([2.2.32], [2.2.33]) and articles ([2.2.34], [2.2.35]) is necessary. It must be stated that no consistent notation for the orbital elements of geostationary satellites has been established in the literature. This leads to confusion, particularly in the case of the inclination vector, as here two categorically different conventions (with axes rotated 90° to each other) are in use.

Finally, the question of **colocating** several geostationary satellites in one control window deserves mention. This can best be achieved by separating the eccentricity and inclination vectors, as was first suggested in 1989 for the TDF-1/-2 and TVSat-2 satellites. The $\boldsymbol{e}/\boldsymbol{i}$ vector separation allows a safe separation of several satellites and is particularly robust against errors that occur when carrying out maneuvers. This is achieved by aligning the relative perigee and line of nodes of each satellite pair so that a separation in the radial and normal direction is always guaranteed. One impressive example is the positioning of up to eight ASTRA satellites at 19.2°E, described in [2.2.37].

Bibliography

[2.2.1] Vallado, D.A. *Fundamentals of Astrodynamics and Applications.* New York: McGraw-Hill, 1997.

[2.2.2] Montenbruck, O., Gill, E. *Satellite Orbits – Models, Methods and Applications.* Heidelberg: Springer Verlag, 2000.

[2.2.3] IERS. http://hpiers.obspm.fr.

[2.2.4] Seidelmann, P.K. (ed.) *Explanatory Supplement to the Astronomical Almanac.* Mill Valley, CA: University Science Books, 1992.

[2.2.5] McCarthy, D.D., Petit, G. IERS Conventions 2003, Observatoire de Paris, Paris, 2003.

[2.2.6] Abramowitz, M., Stegun, I.A. *Handbook of Mathematical Functions.* New York: Dover, 1965.

[2.2.7] Tapley, B., Ries, J., Bettadpur, S. *et al.* GGM02 – An Improved Earth Gravity Field Model from GRACE. *J. Geod.*, **79**, 467–478, 2005.

[2.2.8] Cunningham, L.E. On the Computation of the Spherical Harmonic Terms Needed during the Numerical Integration of the Orbital Motion of an Artificial Satellite. *Celestial Mech.*, **2**, 207–216, 1970.

[2.2.9] Klinkrad, H., Fritsche, B. Orbit and Attitude Perturbations due to Aerodynamics and Radiation Pressure. ESA Workshop on Space Weather, Noordwijk, the Netherlands, November 11–13, 1998, ESTEC.

[2.2.10] Marcos, F.A., Bowman, B.R., Sheehan, R.E. Accuracy of Earth's Thermospheric Neutral Density Models. AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Keystone, Colorado, August 21–24, 2006, AIAA 2006-6167.

[2.2.11] Adhya, S., Ziebart, M., Sibthorpe, A. *et al.* Thermal Force Modeling for Precise Prediction and Determination of Spacecraft Orbits. *Navigation*, **52** (3), 131–144, 2005.

[2.2.12] Beutler, G. *Methods of Celestial Mechanics.* Heidelberg: Springer Verlag, 2004.

[2.2.13] Taff, L.G. *Celestial Mechanics: A Computational Guide for the Practitioner.* New York: John Wiley & Sons, Inc., 1985.

[2.2.14] Kaula, W.M. *Theory of Satellite Geodesy: Applications of Satellites to Geodesy.* Waltham, MA: Blaisdell, 1966. Reprinted New York: Dover.

[2.2.15] Hoots, F.R., Roehrich, R.L. Models for Propagation of NORAD Element Sets. Aerospace Defense Command, United States Air Force, *Project Spacecraft Report No. 3*, December 1980; Second Edition by Kelso, T.S., 1988.

[2.2.16] Hoots, F.R., Schumacher, P.W., Glover, R.A. History of Analytical Orbit Modeling in the US Space Surveillance System. *J. Guid., Navig. Control*, **27** (2), 174–185, 2004.

[2.2.17] Brouwer, D. Solution of the Problem of Artificial Satellite Theory without Drag. *Astron. J.*, **64**, 378–397, 1959.

[2.2.18] Hairer, E., Nørsett, S.P., Wanner, G. *Solving Ordinary Differential Equations I.* Heidelberg: Springer Verlag, 1987.

[2.2.19] Montenbruck, O., Gill, E. State Interpolation for On-board Navigation Systems. *Aerosp. Sci. Technol.*, **5**, 209–220, 2001.

[2.2.20] Shampine, L.F., Gordon, M.K. *Computer Solution of Ordinary Differential Equations.* San Francisco: Freeman, 1975. (German edition, *Computerlösung gewöhnlicher Differentialgleichungen.* Wiesbaden: Vieweg Verlag, 1984.)

[2.2.21] Hartl, Ph. *Fernwirktechnik der Raumfahrt.* Heidelberg: Springer Verlag, 1977.

[2.2.22] Pearlman, M.R., Degnan, J.J., Bosworth, J.M. The International Laser Ranging Service. *Adv. Space Res.*, **30**, (2), 135–143, 2002.

[2.2.23] Jayles, Ch., Vincent, P., Rozo, F. *et al.* DORIS-DIODE: Jason-1 Has a Navigator on Board. *Mar. Geod.*, **27,** 753–771, 2004.

[2.2.24] Misra, P., Enge, P. *Global Positioning System (GPS): Theory and Applications.* Lincoln, MA: Ganga-Jamuna Press, 2001.

[2.2.25] Kroes, R., Montenbruck, O., Bertiger, W. *et al.* Precise GRACE Baseline Determination Using GPS. *GPS Solutions*, **9**, 21–31, 2005.

[2.2.26] Schwarz, H.R. *Numerische Mathematik*, Wiesbaden: B.G. Teubner Verlag, 1988.

[2.2.27] Press, W.H., Flannery, B.P., Teukolsky, S.A. *et al. Numerical Recipes in C – The Art of Scientific Computing,* Second Edition. Cambridge: Cambridge University Press, 1992.

[2.2.28] Bhat, R.S., Frauenholz, R.,B., Cannell, P.E. TOPEX/Poseidon Orbit Maintenance Maneuver Design. AAS/AIAA Astrodynamics Specialist Conference, August 7–10, 1989, Stowe, Vermont: AAS-89-408: 645-676, 1989.

[2.2.29] Frauenholz, R.B., Bhat, R.S., Shapiro, B.E. *et al.* Analysis of the TOPEX/Poseidon Operational Orbit: Observed Variations and Why. *J. Spacecr. Rockets*, **35** (2), 212–224, 1998.

[2.2.30] Guinn, L. Autonomous Navigation Flight Demonstration Results for the New Millennium Program EO-1 Mission. AIAA/AAS Astrodynamics Specialist Conference and Exhibit, August 5–8, 2002, Monterey, California, AIAA 2002-4983.

[2.2.31] Grondin, M., Issler, J.-L., Charmeau, M.-C. *et al.* Autonomous Orbit Control with GPS On Board the DEMETER Spacecraft. NAVITEC'2006, December 11–13, 2006, Noordwijk, the Netherlands.

[2.2.32] Soop, E.M. *Handbook of Geostationary Orbits.* Dordrecht: Kluwer Academic, 1994.

[2.2.33] Sidi, M.J. *Spacecraft Dynamics & Control.* Cambridge: Cambridge University Press, 1997.

[2.2.34] Eckstein, M.C. Geostationary Orbit Control Considering Deterministic Cross Coupling Effects. Congress of the International Astronautical Federation, October 6–12, 1990, Dresden, IAF 90-326, 1990.

[2.2.35] Campan, G., Alby, F., Dufor, F. *et al.* Geostationary Satellite Station Keeping. In *Spaceflight Dynamics, Part 2: 1489–1549.* Toulouse: Cepadues-Editions, 1995.

[2.2.36] Eckstein, M.C., Rajasingh, C.K., Blumer, P. Colocation Strategy and Collision Avoidance for the Geostationary Satellites at 19 Degrees West. International Symposium on Space Flight Dynamics, November 6–10, 1989, Toulouse, CNES, 1989.

[2.2.37] Wauthier, P., Bischops, E., Francken, Ph. *et al.* On the Co-location of Eight Astra Satellites. International Symposium on Space Flight Dynamics, June 26–30, 2000, Biarritz.

2.3 Aerothermodynamics and Reentry

Klaus Hannemann and José Longo

Aerothermodynamics is one of the disciplines in aeronautics and aerospace research which, especially during the last 50 years, has been strongly advanced. Before the flight of the Soviet Sputnik 1 in 1957, representing the first artificial satellite in an Earth orbit, no textbook relating to aerothermodynamics and reentry strategies was available. Since then, however, numerous books have been published and every year several conferences focus on this topic. The discussion of aerothermodynamics and reentry presented here is condensed to an overview of the driving physical and chemical phenomena and provides basic relations and correlations. References allowing a more detailed study of this complex topic are given at the end of this section.

2.3.1 Introduction

Aerothermodynamics can be regarded as an extension of classical aerodynamics of flight vehicles toward higher velocities. While classical **aerodynamics** is mostly concerned with aircraft flying at speeds in the order of the speed of sound, aerothermodynamics analyzes the flight properties at several times the speed of sound. Although no well-defined boundary exists, flows at Mach numbers $M = v/a$ (v and a denote the flight speed and the speed of sound, respectively) of $M \geq 5$ are generally referred to as **hypersonic flows**. During the reentry flight of a **space vehicle** into the **Earth's atmosphere** velocities exceeding 25 times the speed of sound are reached.

With increasing flight velocity, the energy content of the flow increases and the heating of the space vehicle and the surrounding air become more and more important. Consequently, aerodynamics must be extended to account for thermodynamic aspects, thus generating a new discipline called aerothermodynamics. Similar effects as observed during reentry occur in the framework of interplanetary missions when a spacecraft enters the sufficiently dense atmosphere of another planet. One of the substantial differences is the composition of the atmospheres. In comparison to the Earth's atmosphere, which consists of 78% nitrogen, 21% oxygen and 1% noble gases (mostly argon) by volume, the atmosphere of Mars for example is made up of 95% carbon dioxide by volume as well as a 5% fraction composed of nitrogen, argon and, to a smaller extent, oxygen and carbon monoxide. The following discussions are restricted to entry into the Earth's atmosphere.

The reentry of a spacecraft is essentially a braking maneuver to slow down the vehicle from its orbital speed for a safe landing. From an economic as well as a technical point of view, the only way to realize this braking maneuver for vehicles of the size typical for human space missions is to utilize aerodynamic forces and control surfaces. Along its reentry flight path a spacecraft is exposed to different flow environments whose dominating fluid mechanical and chemical processes vary strongly. In addition to the correct treatment of these phenomena, the investigation of the aerothermodynamics of a spacecraft is closely linked to a number of additional disciplines such as flight mechanics or structural mechanics and is thus a multidisciplinary undertaking. The aerothermodynamic properties of a vehicle determine its aerodynamic behavior, and are used to select the flight path and suitable materials for the thermal protection system.

Ground-based testing facilities, including a large variety of different measurement techniques, numerical prediction methods and flight tests, are used as tools to determine aerothermodynamic data.

2.3.2 Global Energy Considerations

Due to the extremely high velocities and the resulting kinetic energy of spacecraft, the aerodynamic braking maneuver in the atmosphere initially appears to be almost impossible. A simple estimation of the orbital velocity results from equating the centrifugal force with the gravitational force. The velocity in a low Earth orbit is then given by (see also Section 2.2)

$$v_u = \sqrt{g_h \cdot (h_u + R_\oplus)} \sim \sqrt{g_h \cdot R_\oplus} \qquad (2.3.1)$$

where:

v_u = speed in orbit [m/s],
g_h = acceleration due to gravity in orbit [m/s^2],
h_u = flight altitude [m],
$R_\oplus$ = radius of the Earth [m].

Based on this relation, the velocity of the International Space Station (ISS) at a flight altitude of $h_u \approx 400$ km is $v_u \approx 7.9$ km/s. Were the complete amount of kinetic energy e_{kin} during reentry from the ISS to be absorbed by the thermal protection system of the spacecraft, this would result in a maximum integral specific heat quantity of

$$\overline{q}_w \approx e_{kin} = v_u^2/2 \approx 31 \text{ MJ/kg} \qquad (2.3.2)$$

A comparison of this heat quantity to that necessary to melt or vaporize some of the elements which are typically used for spacecraft (Table 2.3.1) indicates that none of the elements shown would withstand such high thermal loads. Regarding pure carbon, it should be noted that while it sublimes at a rather high temperature, it would burn in the atmosphere.

An estimation of the temperature in the stagnation region of a spacecraft, assuming that the air behaves like a calorically perfect gas, results in

Table 2.3.1: *Energy necessary to melt or vaporize different elements.*

Element	Heat of fusion [MJ/kg]	Heat of vaporization [MJ/kg]	Melting point [K]
Aluminum	0.40	10.9	933
Tungsten	0.19	4.5	3695
Titanium	0.32	8.8	1941
Molybdenum	0.41	6.2	2896
Silicon	1.8	13.6	1683
Beryllium	1.35	32.44	1551
Carbon	–	59.5	3820

$$h \approx v_u^2/2 \approx 31\ \text{MJ/kg};\quad T = h/c_p \approx 31\,000\ \text{K} \qquad (2.3.3)$$

where:

h = specific enthalpy [J/kg],
T = temperature [K],
c_p = specific heat capacity at constant pressure [J/(kg K)].

Comparing the specific enthalpy of the air in the stagnation region to the energy required to dissociate the main components of air (15 MJ/kg for oxygen and 34 MJ/kg for nitrogen) shows that the composition of air at the above-mentioned conditions changes due to the resulting dissociation and ionization of molecules and atoms. Because of the high-temperature environment, the translational energy of the air particles reaches a level causing the energy exchange during particle collisions to be high enough to generate so-called high-temperature effects. These will be discussed in more detail in the following section.

The global consequence of these endothermic processes is a strong reduction of the gas temperature in the stagnation region. This discussion reveals already that in reality the complete energy which is converted into heat during the braking maneuver is not absorbed by the structure of the spacecraft. A large fraction of the energy is absorbed by the air in the vicinity of the vehicle and subsequently carried away by the flow. In a first approximation, the fraction of the kinetic energy which is involved in the heating of the vehicle's surface due to friction is [2.3.1]

$$Q_w = 1/2 \cdot (C_f / C_d) \cdot 1/2 \cdot m \cdot v_u^2 \qquad (2.3.4)$$

where:

Q_w = integral heat quantity [J],
C_f = averaged vehicle skin friction coefficient [dimensionless],
C_d = drag coefficient [dimensionless],
m = vehicle mass [kg].

This relationship shows that the integral heat quantity which is emitted to the surface depends on the kinetic energy of the vehicle in orbit and the ratio of viscous drag to total drag, consisting of pressure drag and viscous drag. Since the viscous drag of turbulent boundary layers is approximately 2–3 times larger than that of laminar boundary layers, the laminar–turbulent transition should be delayed as much as possible during the flight path in order to reduce viscous drag. Further, it is clear that vehicles with a pressure drag predominating the viscous drag experience a smaller integral heat load.

The dependence of the integral heat quantity on the viscous drag appearing in the above estimate is used here as an opportunity to additionally refer to a general relationship of practical importance. In 1875, Reynolds identified an analogy between the transport of momentum, energy and mass. This analogy states that for boundary layer flows of fluids with $Pr = Sc = 1$ and zero pressure gradient along the surface, the Stanton number, representing the nondimensionalized wall heat flux $q_{w,gas}$ (see also Section 2.3.4), can be determined by

$$St = C_f / 2,$$

where:

Pr = **Prandtl number**, $Pr = \mu \cdot c_p/k$ [dimensionless],
Sc = **Schmidt number**, $Sc = \mu / (\rho \cdot D)$ [dimensionless],
μ = dynamic viscosity [N s/m^2],
κ = thermal conductivity [W/(m · K)],
ρ = density [kg/m^3],
D = diffusion coefficient [m^2/s],
St = Stanton number, $St = q_{w,gas}/(\rho_\infty \cdot V_\infty \cdot (h_0 - h_w))$ [dimensionless],
$q_{w,gas}$ = wall heat flux [W/m^2],
h_0 = total specific enthalpy [J/kg],
h_w = wall specific enthalpy [J/kg].

Extensions of this relationship for flows which do not meet the above-mentioned requirements are given, for example, in [2.3.2].

2.3.3 Fluid Mechanical and Chemical Phenomena during Reentry

The overview in Figure 2.3.1 summarizes the most important fluid mechanical and chemical processes which occur during reentry of a spacecraft in the Earth's atmosphere. Two different reentry corridors are considered here, one from a low orbit at an altitude of approximately 400 km and one from an interplanetary or Moon mission. Related to the different

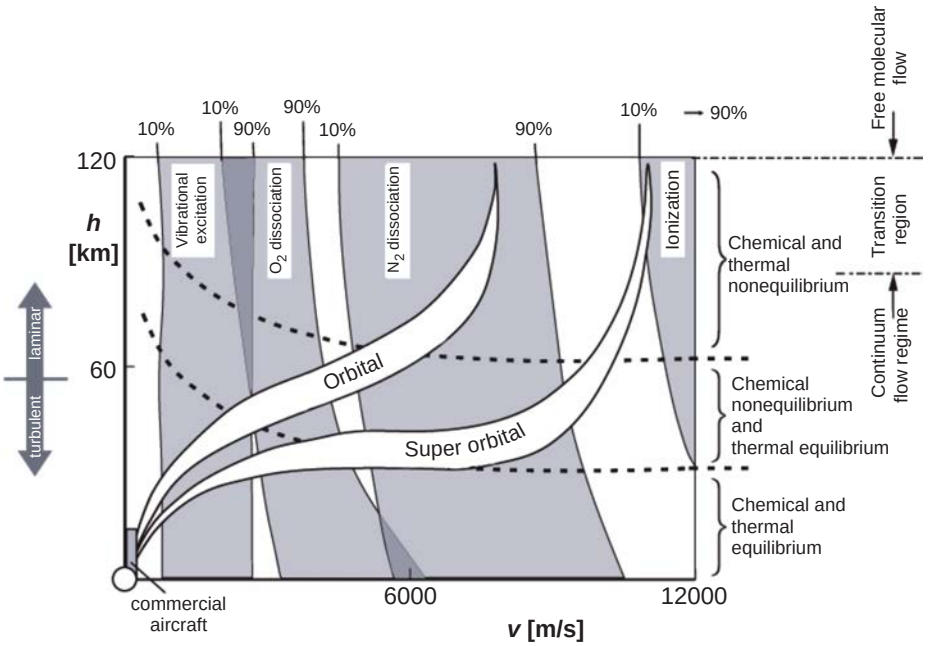


Figure 2.3.1: *Overview of fluid mechanical and chemical processes occurring along reentry flight paths.*

mission profiles, a distinction is made between orbital and superorbital reentry. In addition to the different velocities at the beginning of the reentry phase, the two flight corridors shown in Figure 2.3.1 also differ because for the **orbital reentry** a winged vehicle was selected and for the **superorbital reentry** a capsule was selected as reference configurations. The different aerodynamic properties of these configurations will be discussed in more detail in the following sections.

The transitions between regimes of different physical and chemical properties shown in Figure 2.3.1 depend on the chosen reference length and vary when different configurations are considered. Further, the boundaries shown have only symbolic character. In reality, no clear-cut dividing lines exist between the different regimes.

During the first part of reentry – after leaving orbit and before the braking maneuver starts in the upper atmosphere – the spacecraft is flying in the regime of free molecular flow. While proceeding further along the reentry flow path, the spacecraft reaches a transition region and subsequently the continuum flow regime. These different regimes are characterized by the **Knudsen number**, $Kn = \lambda / L$, where λ is the mean free path [m] and L the characteristic length [m]; of the vehicle, both given in [m]. A flow is considered to be a free molecular flow if the distance λ traveled by an air particle between two consecutive collisions is several times larger than this characteristic length. Due to the decreasing density of the atmosphere with altitude, the orbiter of the Space Shuttle is located in this regime in flight altitudes above 150 km. In the Knudsen number regime $Kn \leq 0.01$, the flow is considered to be a continuum flow; that is, the mean free path is significantly smaller than the characteristic length. Due to the different number of particles which hit a spacecraft in the different regimes, the aerodynamic properties change and the set of equations required for the mathematical description of the flow needs to be adapted accordingly (e.g., [2.3.2]; see also Section 2.3.7).

During the aerothermodynamic design process of a hypersonic vehicle, viscous effects are a key factor. The heat flux at the surface of blunt reentry configurations is of particular practical importance and will subsequently be addressed in more detail. The parameter which characterizes the influence of viscous effects on a flow field is the Reynolds number, $Re = \rho \cdot v \cdot L / \mu$.

2

As indicated in Figure 2.3.2, the **Reynolds number** varies over several orders of magnitude along a reentry trajectory. In high-altitude flight the wall boundary layer of a reentry vehicle is initially laminar. After exceeding a critical Reynolds number, the **transition from a laminar to a turbulent boundary layer** takes place. This process is linked with an increase of the skin friction and, according to the Reynolds analogy, an increase of the wall heat flux.

The high-energy content of reentry flows leads to strong heating of the air in the vicinity of a spacecraft. The resulting high-temperature effects occurring behind the bow shock wave developing in front of a reentry vehicle are schematically shown in Figure 2.3.3. In the nose region of the configuration the bow shock has its highest strength and the temperature increase across the shock reaches a maximum. Depending on the temperature level behind the shock wave, the vibrational degrees of freedom of the air molecules are excited (1) and dissociation reactions of oxygen and nitrogen molecules may occur (2). With increasing temperature in the bow shock region, gas radiation (3) adds to spacecraft heating. Further, ionization of the air constituents (4) occurs, which leads to the so-called "black-out" effect, namely a disruption of radio communications between the ground control station and the spacecraft. The high-temperature effects described here are enabled by energy transfer from the translational energy stored in the random motion of the air particles, which is increased by the gas heating, to other forms of energy. Because this energy transfer is realized by air particle collisions, it requires a certain time period to develop. The time required to reach an equilibrium condition, which, for example, is defined by the local temperature and density, is called the relaxation time. Whether this condition can be reached depends on the Damköhler number $D = t_S / t_R$.

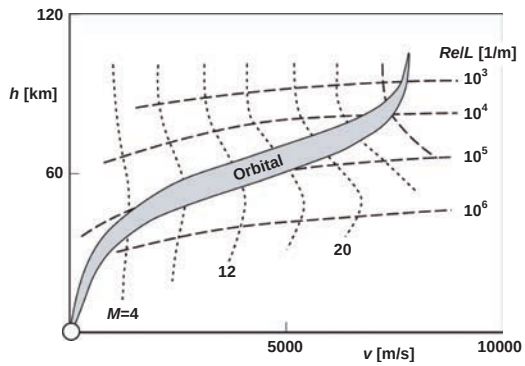


Figure 2.3.2: *Variation of Mach and Reynolds numbers (based on a reference length of 1 m) along a typical reentry corridor.*

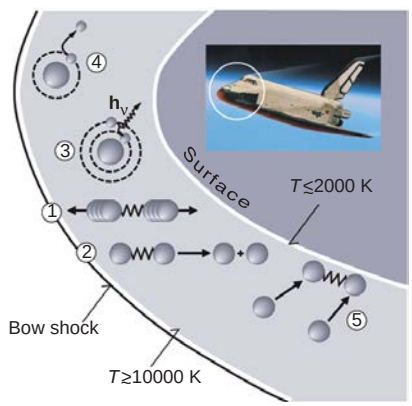


Figure 2.3.3: *Overview of the fundamental high-temperature effects occurring in the shock layer in front of a reentry vehicle.*

In the case $D \gg 1$, the characteristic time scale of the flow, that is the residence time t_S of a particle in a certain flow environment, is large compared to the relaxation time t_R and the flow is in equilibrium. The other limiting case $D \ll 1$ characterizes a frozen flow because no relaxation processes can advance and $D \approx 1$ describes a flow in nonequilibrium.

From Figure 2.3.1 it can be seen that during reentry regimes occur in which the flow can be in nonequilibrium or equilibrium regarding the thermal excitation (of internal degrees of freedom such as vibration) and chemical relaxation.

The influence of **chemical reactions** on the flow topology in the case of a reentry capsule is demonstrated in Figure 2.3.4. The depicted Mach number distribution in the symmetry plane and the skin friction lines result from numerical computations of the three-dimensional viscous flow past the capsule configuration within the framework of the European flight experiment "Atmospheric Reentry Demonstrator" (ARD). The reentry flight experiment was performed in 1998 using the Ariane 5 launcher [2.3.3]. In order to clarify the influence of chemical reactions, the flow past the ARD capsule was computed for a flight Mach number $M = 24$ at an altitude of 65 km assuming

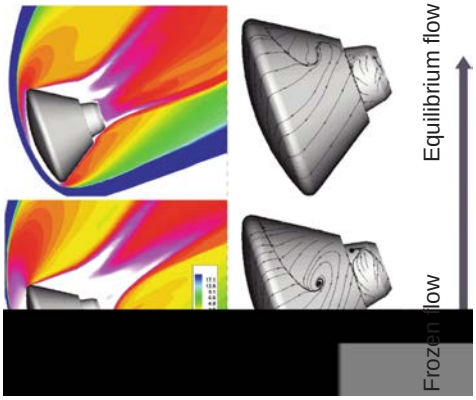


Figure 2.3.4: *Influence of high-temperature relaxation effects on the topology of the flow past a capsule configuration at M = 24 and angle of attack of α = 20°.*

both limiting cases: a flow in equilibrium and a frozen flow [2.3.4]. The difference in flow topology between both solutions is self-evident. A characteristic feature of high enthalpy hypersonic flows past blunt bodies is that compared to the frozen flow the shock standoff distance is significantly reduced due to chemical reactions and the associated increase in shock layer density. This results in a lower surface pressure level and hence lower forces and moments in the case of the frozen flow. For the example shown in Figure 2.3.4 the aerodynamic coefficients differ by 50% for the pitching moment, 20% for the lift and 5% for the drag.

Depending on the properties of the vehicle surface, the energy stored in the form of dissociation energy in the atoms can be released again at the surface, causing additional heating. This would be the case for a **catalytic surface**, that is a surface at which recombination reactions take place (see (5) in Figure 2.3.3) [2.3.5]. For a **noncatalytic surface** this additional heating does not occur. As a consequence, for **thermal protection systems** for spacecraft, materials are selected which preferably have noncatalytic properties [2.3.6]. Examples are silicon carbide (SiC), which is highly temperature resistant (sublimation temperature ≈ 2900 K) and for a ceramic material highly heat conductive, or the composite material carbon–silicon carbide (C–SiC).

The results of computations taking into account chemical nonequilibrium in the flow past the hyperboloid depicted in Figure 2.3.5 suggest that compared to a fully catalytic wall the heat flux in the stagnation point of a reentry configuration can be 50% lower for noncatalytic walls [2.3.7]. The geometry of the hyperboloid has been chosen such that for an axisymmetric free stream, the flow conditions along the windward symmetry line of the **Space Shuttle Orbiter** are approximately reproduced. The numerical and flight data correspond to a Mach number of $M = 27$ at an altitude of approximately 86 km. Comparison of measured and computed data indicates that the properties of the thermal protection system of the Space Shuttle Orbiter change at approximately $x = 3$ m from partially catalytic to fully catalytic. In this region, however, the absolute value of the wall heat flux is already strongly reduced.

Further discussions of the phenomena addressed in this section can be found, for example, in [2.3.2], [2.3.6], [2.3.8] or [2.3.9].

2.3.4 Heat Flux Balance

The heat flux balance on the surface of a vehicle flying at hypersonic speed can be described as follows:

$$q_{\text{gas,cond}} + q_{\text{gas,diff}} + q_{\text{gas,rad}} - q_{\text{w,rad}} = q_{\text{w}} \qquad (2.3.5)$$

The heat flux into the surface structure of a vehicle q_{w} is composed of contributions resulting from **heat**

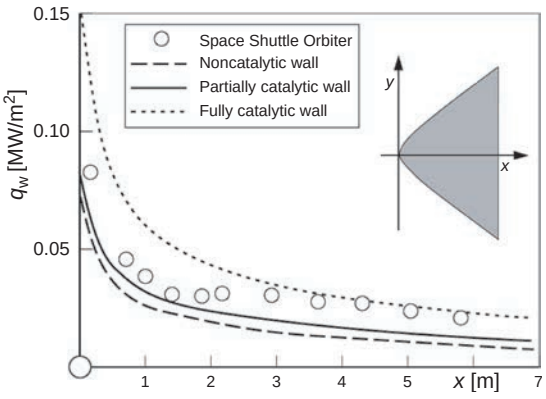


Figure 2.3.5: *Influence of the wall catalysis on the heat flux of a generic reentry configuration at M = 27 at 86 km altitude.*

conduction and diffusion processes in the gas $q_{w,gas} = q_{gas,cond} + q_{gas,diff}$ (see also Section 2.3.2), **gas radiation** $q_{gas,rad}$ and heat radiation of the hot surface $q_{w,rad}$. The surface heat flux due to conduction in the gas, $q_{gas,cond} = -\kappa \cdot (\partial T / \partial y)_{W,}$ where y is the wall normal coordinate in meters, is given by Fourier's law. Energy transport due to diffusion in chemically reacting flows is caused by the emerging gradients of species concentrations. The contribution of diffusion to the wall heat flux can be expressed by

$$q_{gas,diff} = -\rho \cdot \sum_i D_{im} \cdot h_i \cdot (\partial c_i / \partial y)_w$$

where:

D_{im}= diffusion coefficient of species i [m^2/s],
h_i = specific enthalpy of species i [J/kg],
c_i = mass fraction of species i [$kg_i/kg_{mixture}$].

This expression is based on the assumption that the diffusive mass flux of each individual species is given by Fick's law. The diffusion coefficient D_{im} describes the diffusion process of species i into the gas mixture consisting of the remaining species of the multicomponent mixture [2.3.2]. The expressions for the heat conduction and the diffusion in the gas are based here on the simplifying assumption that the processes can be regarded as locally one dimensional.

The part of a reentry vehicle which generally experiences the highest thermal loads is the stagnation point region of the blunt nose with radius R. Therefore, relationships have been derived which allow estimations of the wall heat flux contribution caused by the gas. Assuming that the region between the bow shock wave and the boundary layer edge is in equilibrium, Fay and Riddell have published correlations for the wall heat flux contributions resulting from heat conduction and diffusion in the gas (see, e.g., [2.3.2]). The case of a frozen boundary layer and a noncatalytic wall yields

$$q_{gas,cond} + q_{gas,diff} = 0.76 \cdot Pr^{-0.6} \cdot (\rho_e \mu_e)^{0.4} \cdot (\rho_w \mu_w)^{0.1} \sqrt{1/R \cdot (2 (p_e - p_\infty)/\rho_e)^{0.5}} \cdot (1 - h_D / h_{0e}) \quad (2.3.6)$$

where:

R = nose radius [m],
p = pressure [Pa],
ρ = density [kg/m^3],
μ = dynamic viscosity [$N \cdot s/m^2$],
h_D = mixture heat of formation, $h_D = \sum_i c_{ie} \cdot \Delta h_{fi}$ [J/kg],
Δh_{fi} = heat of formation of species i [J/kg],
h_{0e} = specific total enthalpy at boundary layer edge [J/kg].

Here, the indices w, e and ∞ denote the wall, the boundary layer edge and the free stream, respectively, and h_D is the energy available at the boundary layer edge which is stored as heat of formation. Further correlations of this form for different gas and surface conditions can be found in [2.3.2]. A simple engineering approximation of the **surface heat flux** caused by the gas is given by Detra and Hidalgo (see [2.3.6]):

$$q_{w,gas} = 11.03 \cdot 10^7 \cdot 1/\sqrt{R} \cdot (\rho_\infty / \rho_{SL})^{0.5} \cdot (v_\infty / v_u)^{3.15} \quad (2.3.7)$$

where:

$q_{w,gas}$ = surface heat flux caused by the gas [W/m^2],
ρ_{SL} = density at sea level [kg/m^3].

An important aspect which can be gathered from the latter two correlations is that the wall heat flux caused by conduction and diffusion in the gas is proportional to $1/\sqrt{R}$ and that it can thus be reduced by preferably choosing a large nose radius.

A mechanism which further contributes to the surface heating of a reentry vehicle is gas radiation. For an optically thin gas, that is a gas which solely emits and does not absorb radiation, and assuming that the distance of the bow shock which develops in front of the spacecraft is small compared to the nose radius, the wall heat flux contribution is given by

$$q_{gas,rad} = E_{rad} / 2 \cdot R \cdot (\rho_\infty / \rho_{BS}) \quad (2.3.8)$$

where:
E_{rad} = power density of radiation [$J/(s\ m^3)$],
ρ_{BS} = postshock density [kg/m^3].

In this relationship, E_{rad} is the radiation energy emitted by the gas per unit time and unit volume which contributes to the wall heat flux at the stagnation point

of a vehicle. An engineering estimate of the wall heat flux due to gas radiation is stated in [2.3.6]:

$$q_{\text{gas,rad}} = 7.9 \cdot 10^{11} \cdot R \cdot (\rho_\infty / \rho_{\text{SL}})^{1.5} \cdot (v_\infty / 10^4)^{12.5} \quad (2.3.9)$$

where $q_{\text{gas,rad}}$ is the wall heat flux due to gas radiation, in W/m^2.

These correlations reveal that the influence of gas radiation is directly proportional to the nose radius of the vehicle. With increasing nose radius, the bow shock stand-off distance increases and consequently the volume of the radiating gas in the shock layer is increased, resulting in a higher surface heat load caused by radiation. Hence, in the framework of the design process of a reentry vehicle, the reduction of the wall heat flux due to conduction and diffusion in the gas and gas radiation is governed by contradictory criteria. The level of both heat flux sources and thus the determination of which contribution dominates the design is governed by the emissivity of the gas and the velocity of the vehicle. During reentry from low orbit, the governing parameters are such that at an altitude of 100 km the wall heat flux based on gas radiation is less than 0.1% of the heat flux caused by conduction and diffusion in the gas. However, during reentry in the framework of an interplanetary mission ($v_u > 11200$ m/s) the heat flux based on gas radiation is significant. During the Apollo return missions from the Moon, the gas radiation contribution amounted to approximately half of that caused by conduction and diffusion. A schematic of these interrelations is given in Figure 2.3.6 [2.3.1].

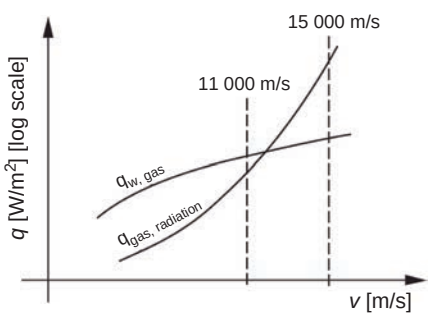


Figure 2.3.6: *Wall heat flux contributions during reentry caused by heat conduction and diffusion and gas radiation.*

According to the **Stefan–Boltzmann law**, the heat flux which is radiated from the hot wall of a spacecraft can be determined by $q_{\text{w,rad}} = \varepsilon \cdot \sigma \cdot T_{\text{w}}^4$, where ε is the emissivity of a surface, $0 \le \varepsilon \le 1$, and σ the **Stefan–Boltzmann constant**, $5.67 \cdot 10^{-8}$ W/(m^2 K^4). For silicon carbide, for example, the emissivity is $\varepsilon \approx 0.9$. Heat radiation of the wall is a very efficient passive cooling system. During reentry, approximately 80% of the wall heat flux generated by the braking maneuver of the vehicle is radiated from the surface and subsequently carried away by the flow.

In addition to **radiation cooling**, further possibilities exist to passively cool spacecraft. These are qualitatively discussed in the following. The heat sink is a technique in which absorbed heat is stored in a body without causing material damage. In this context heat storage vessels in the form of, for instance, air chambers or water reservoirs are applied. Ablation cooling is used in the case of very high thermal loads. Here, the heat generated by the braking maneuver is dissipated by endothermic decomposition, melting and vaporization of the heat shield. The layer generated by the molten and vaporized surface material acts as a thermal protection system and avoids high heat fluxes onto the spacecraft. Materials used for **ablation cooling** are characterized by high melting points, high heats of fusion, high heats of vaporization and low heat conductivity. Examples of ablative thermal protection system materials are fiber-reinforced plastic or graphite. Ablation cooling results in the consumption of the thermal protection system material. Due to the extremely high thermal loads occurring during interplanetary missions, from a technical point of view this kind of cooling is the only feasible solution. During entry into the atmosphere of Jupiter, the capsule of the Galileo spacecraft reached a velocity of approximately $50–60 \cdot 10^3$ m/s whereby during the heating process the ablative carbon fiber-reinforced thermal protection system reached temperatures of up to 8800 K.

In order to select a thermal protection system for a spacecraft it is necessary to know the planned flight trajectory (Section 2.3.5). As shown in Figure 2.3.7, during reentry of two different types of spacecraft a similar integral local heat flux into the structure can occur. However, both types experience quite different maximum loads. The integral local heat flux is given

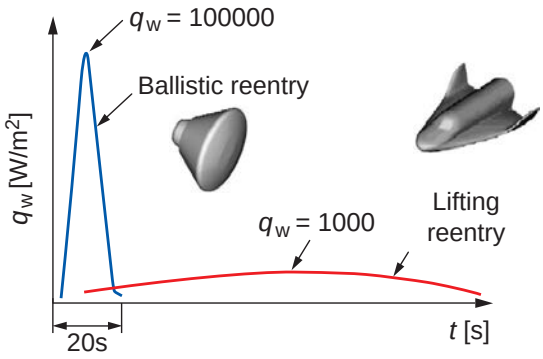


Figure 2.3.7: *Sketch of the temporal development of the maximum local heat fluxes experienced by a capsule in a ballistic trajectory and by a winged vehicle in a lifting trajectory.*

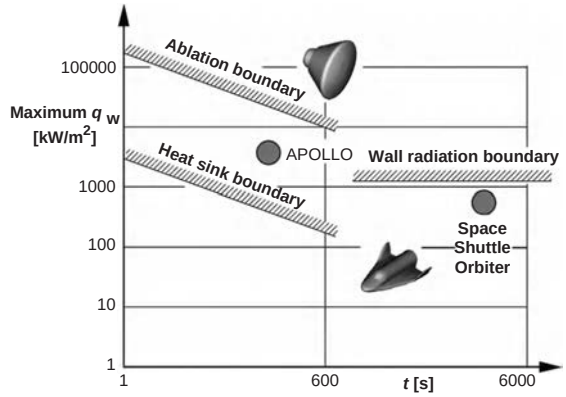


Figure 2.3.8: *Limit of applicability of different passive cooling concepts depending on the duration of the flight and the maximum heat flux.*

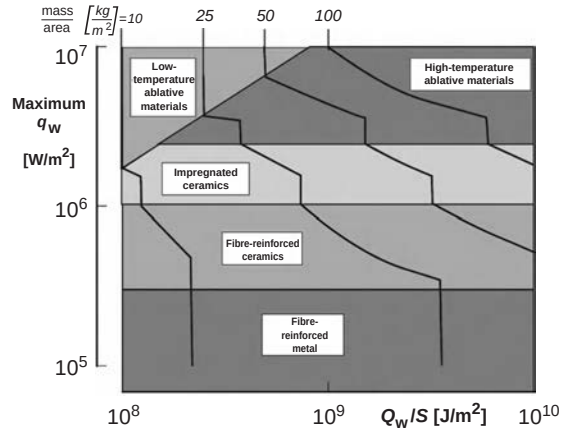


Figure 2.3.9: *Weight of thermal protection system utilizing different concepts depending on the maximum heat flux and the accumulated heat quantity per unit surface area [2.3.10].*

by the area below the sketched temporal development of the maximum local heat flux. Along the ballistic trajectories of the capsule configurations, which generate almost no lift, extremely high heat fluxes occur over a short period of time. In contrast, along the flight path of lift-generating space planes, significantly lower values are present for a longer period of time.

In Figure 2.3.8, the applicability limits for the three passive cooling methods discussed here are given, depending on the duration of the reentry flight and the maximum heat flux. The dimensions of the thermal protection system increase with increasing accumulated heat quantity. Consequently, as depicted in Figure 2.3.9, the weight of the thermal protection system increases. This is a critical design parameter for a spacecraft. For a vehicle like the Space Shuttle Orbiter, which was designed for reentry missions from low orbits, such as from the ISS, the weight of the thermal protection system amounts to approximately half of the overall payload. The weight of the thermal protection system of the capsule of the Galileo space probe which was designed for entry into Jupiter's atmosphere reached approximately half of the vehicle weight.

2.3.5 Reentry Trajectory

The reentry path depends on the properties of the atmosphere. These are the basis for each trajectory calculation and are available in the form of standard atmospheres [2.3.1].

According to Newton's laws the following equations of motion for a reentry vehicle can be derived (see, e.g., [2.3.1], [2.3.11]):

$$W \cdot \sin\theta - D = m \cdot \mathrm{d}v/\mathrm{d}t \qquad (2.3.10)$$

$$L - W \cdot \cos\theta = -m \cdot v^2/R \qquad (2.3.11)$$

where:

W = weight, $W = m \cdot g$ [N],
D = drag [N],

L = lift [N],
θ = flight trajectory angle [°],
R = radius of flight path curvature [m].

The first equation describes the balance of forces along the flight path and the second equation the one normal to the flight path. Substituting the definition of drag

$$D = 1/2 \cdot \rho \cdot v^2 \cdot C_d \cdot S$$

where S is the reference area in meters, the equation of motion along the flight path can be written as

$$-1/g \cdot dv/dt = (W/(C_d \cdot S))^{-1} \cdot \rho v^2/2 - \sin\theta \quad (2.3.12)$$

Replacing the lift by $L = 1/2 \cdot \rho \cdot V^2 \cdot C_l \cdot S$, where C_l is the dimensionless lift coefficient, in the equation of motion normal to the flight path results in

$$\cos\theta - 1/g \cdot v^2/R = (W/(C_d \cdot S))^{-1} \cdot C_l/C_d \cdot \rho v^2/2 \quad (2.3.13)$$

The solutions of these equations are influenced by the **ballistic coefficient** $W/(C_d \cdot S)$ and the **lift/drag ratio** C_l/C_d, which in turn determines the aerodynamic performance of a configuration. For a purely ballistic flight path, that is $C_l = 0$, and a given flight trajectory angle, the ballistic coefficient is the only determining factor. Lift-generating configurations are characterized by $C_l/C_d \approx 1$ and ballistic configurations by $C_l/C_d << 1$. From the equations given above, the complete flight trajectory, that is the velocity as a function of the flight altitude, the maximum deceleration, the altitude at which the maximum deceleration occurs, etc., can be determined [2.3.1], [2.3.11].

In Figure 2.3.10, the qualitative variation of some relevant flow parameters along a reentry path is plotted. In addition to the density and the velocity, the quantities $\rho \cdot v^2$ and $\rho \cdot v^3$ are given here, which are of great importance for aerodynamic and heat flux considerations. From the definition of aerodynamic drag it is apparent that it scales with $\rho \cdot v^2$. The nondimensionalized heat flux caused by heat conduction and diffusion in the gas is given by the **Stanton number** (see Section 2.3.2). Using the common hypersonic approximation for the specific total enthalpy $h_0 \approx v_\infty^2/2$, and $h_0 \gg h_w$, yields $q_{w,gas} = 1/2 \cdot \rho_\infty \cdot v_\infty^3 St$. Note that the heat flux scales with $\rho \cdot v^3$. The plots in Figure 2.3.10 clearly show that both quantities, namely the drag which causes a load on the vehicle structure as well as the heat flux, pass through a maximum during reentry. During the design of reentry vehicles it is important to ensure that the maxima of structural and thermal loads do not appear at the same flight altitude.

For reasons of economy, the reentry flights realized up to now are based on vehicles without propulsion systems. The possible reentry corridor of a vehicle is limited by two trajectories, namely the undershoot and the overshoot trajectory. The overshoot trajectory is typically defined as the one with the shallowest allowable angle of entry at the atmospheric interface required to prevent skip-off. The undershoot trajectory is defined by the steepest allowable trajectory.

Two basically different kinds of reentry trajectories used so far can be distinguished. On the one hand there is the ballistic flight path as employed by the Apollo capsules and which is still flown by the Soyuz capsules. As already mentioned above, a ballistic flight path is defined by the spacecraft generating no lift. Strictly speaking, this does not completely apply to the capsules mentioned here; however, the lift generated

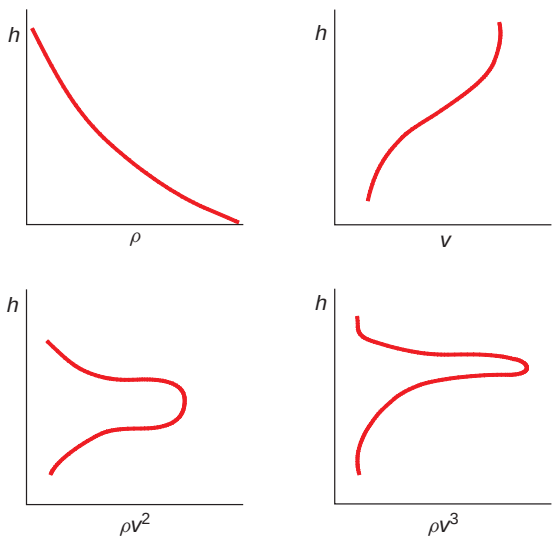


Figure 2.3.10: *Typical variation of different flow parameters along a reentry path.*

is so slight that the flown trajectory is very close to a ballistic one. Ballistic trajectories are in general very steep (undershoot trajectory). They are determined by the maximum acceptable deceleration for human missions and the tolerable thermal and mechanical loads. In the framework of human missions, accelerations during the braking maneuver of 10 times the acceleration of gravity are regarded as an upper limit. Space transportation systems which fly nearly ballistic trajectories utilize blunt capsule configurations, ablative thermal protection systems and parachutes for the landing approach. A disadvantage of this type of trajectory is that the cross-range capability, namely the range normal to the orbiting plane, is limited. On the other hand, lift-generating reentry configurations allow for a significantly less steep trajectory, which simultaneously leads to a reduction of the maximum heat fluxes (see Figure 2.3.7). This type of reentry trajectory is flown, for example, by the Space Shuttle Orbiter. Further advantages of a lifting trajectory are a larger cross-range capability and improved landing accuracy.

As mentioned earlier, an important parameter for the characterization of reentry trajectories is the lift/drag ratio. In addition to the flight range and cross-range capability, the selection of the possible point in time to commence the deorbit maneuver, that is to begin the reentry phase, is determined by this parameter. Particularly from a safety point of view, this is an important issue. In Figure 2.3.11, the waiting time between two consecutive possible deorbit maneuvers preparatory to reentry and landing at a predefined location on Earth is given as a function of the lift/drag ratio and the orbit inclination. The waiting time is drastically reduced with increasing lift/drag ratio. Spacecraft with low lift/drag ratio – primarily capsule configurations – are able to return to a defined landing point only once in 24 hours. With a lift/drag ratio of $C_l / C_d \sim 2$, the Space Shuttle Orbiter has several opportunities to return to Earth in the course of a day. In order to be able to reach any point on Earth from a given orbit, the cross-range based on the radius of the orbit must be at least $\pi/2$. To achieve this, a configuration would need a ratio of $C_l / C_d = 3.5$ [2.3.12].

The criteria defining the boundaries of the reentry corridor of a lift-generating configuration are

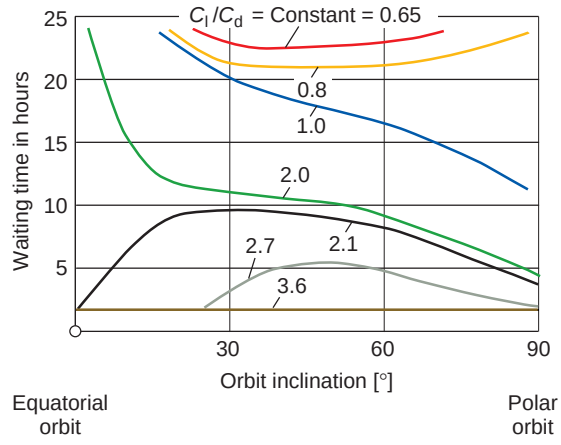


Figure 2.3.11: *Waiting time between two consecutive possible deorbit maneuvers as a function of the lift/drag ratio and the orbit inclination.*

depicted in Figure 2.3.12. The upper boundary is determined by the aerodynamic properties of the vehicle, that is the lift/drag ratio. The lower boundary is specified by the maximum heat flux at high flight velocities, the maximum acceleration load for the crew during the breaking maneuver at medium velocities, and the structural load limit of the vehicle at low velocities.

Reentry vehicles with large lift/drag ratios are difficult to realize because they would exhibit a low volumetric efficiency (ratio of surface to volume). From Figure 2.3.13 it is clear that configurations such as capsules or sphere/cone/cylinders with high volumetric efficiencies are characterized by low lift/drag ratios. In addition, Figure 2.3.13 suggests that for reentry configurations lift/drag ratios above 3 are difficult to achieve.

A further important facet of the design of hypersonic configurations is of course aerodynamic stability. Regarding this aspect, reference is made here to the corresponding special literature (e.g., [2.3.1], [2.3.11]) for a more detailed discussion.

2.3.6 Aerodynamic Considerations

Some relationships for the determination of aerodynamic loads are presented here based on so-called

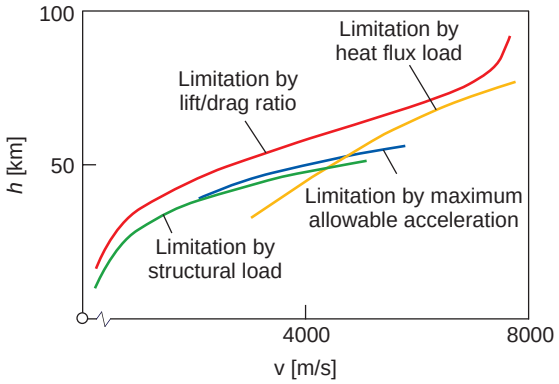


Figure 2.3.12: *Boundaries defining the reentry corridor of a lift-generating reentry vehicle.*

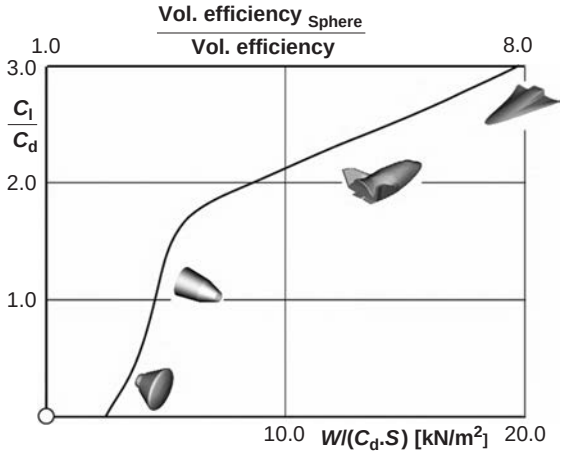


Figure 2.3.13: *Lift/drag ratio and volumetric efficiency of different configurations based on that of a sphere, as a function of the ballistic coefficient.*

surface inclination methods. The application of these methods is restricted to the determination of aerodynamic parameters of spacecraft flying at hypersonic speeds and 10°–50° angles of attack. This is the most important regime for reentry configurations which are characterized by a lift/drag ratio of C_l/C_d = 1–3. For significantly lower velocities and angles of attack, the determination of aerodynamic forces is more complex and cannot be addressed in this section.

The aerodynamic forces and moments acting on a vehicle primarily depend on its shape. For hypersonic flight, the resulting pressure distribution on a body can be assessed by simple engineering models. As already mentioned, flows at high Mach number are characterized by the total specific enthalpy's being approximately equal to the specific kinetic energy and the static enthalpy's being negligible. When this is assumed, the force exerted by the flow on a surface is caused solely by the kinetic energy of the flow particles. A corresponding mathematical flow model was developed by Isaac Newton (see, e.g., [2.3.2]), albeit for flows at low Mach number, where these assumptions do not apply. It states that the flow particles lose their momentum normal to a surface when impinging on it and that they subsequently move in a surface tangential direction. With this assumption, the nondimensionalized difference between the surface pressure and the static pressure of the oncoming flow can be expressed by the relationship (δ represents the angle between a surface and the free stream velocity)

$$C_p = k \cdot \sin^2 \delta \tag{2.3.14}$$

In Newton's original theory $k = 2$ was used for the form factor. This is only valid for the limiting case of an infinitely high Mach number and a ratio of specific heats of $\gamma = 1$. Considering a blunt reentry configuration, the stagnation point of the vehicle is located at $\delta = 90°$. Here the maximum value of C_p occurs, namely $C_p = C_{pmax}$. The pressure at the stagnation point can be obtained by determining the stagnation pressure behind a normal shock wave, the so-called Pitot pressure. Consequently, the form factor for the modified Newton method is $k = C_{pmax}$, with

$$\begin{aligned} C_{pmax} &= (p_{t2} - p_\infty) / (0.5 \cdot \rho_\infty \cdot v_\infty^2) \\ &= 2/(\gamma M_\infty^2)\,(p_{t2} / p_\infty - 1) \end{aligned} \tag{2.3.15}$$

where p_{t2} is the Pitot pressure in pascals and γ the adiabatic coefficient, $\gamma = c_p/c_v$. Therefore, the form factor depends on the flight Mach number and the ratio of specific heats at constant pressure c_p and constant volume c_v. This dependence is depicted in Figure 2.3.14.

In Figure 2.3.15, the pressure coefficient of a cone with a half cone angle of 15° resulting from an exact solution of the Euler equation and the **Newton method** is shown as a function of the **Mach number**. Both solutions converge with increasing Mach number. This

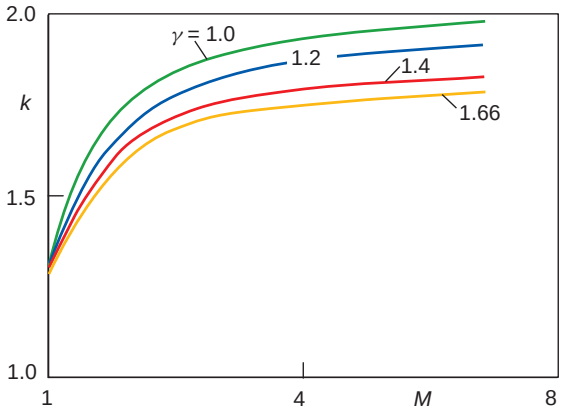


Figure 2.3.14: *Form factor k as a function of the free stream Mach number and the adiabatic coefficient γ.*

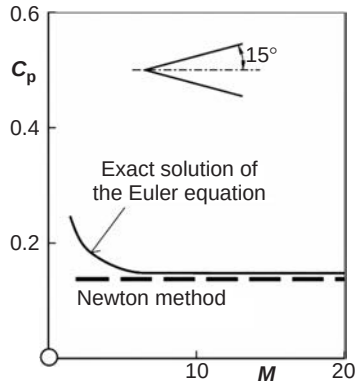


Figure 2.3.15: *Variation of the pressure coefficient of a cone as function of Mach number resulting from an exact solution of the Euler equation and the Newton method.*

result indicates that the Newton method is applicable for hypersonic flows. A second aspect which becomes apparent from Figure 2.3.15 is the Mach number independence principle of hypersonic flows. As long as the ratio of specific heats γ of the gas remains constant, the pressure coefficient is independent of the Mach number from approximately $M \approx 7$ [2.3.13]. As a consequence, lift and pressure drag are also independent of the Mach number.

The method to determine the surface pressure discussed so far can only be applied in the continuum regime of the trajectory, that is in a flight altitude range below approximately 90 km, where relatively high dynamic pressures act on a vehicle. The determination of aerodynamic coefficients in the rarefied flow regime is only possible with numerical methods such as the **direct simulation Monte Carlo (DSMC)** method (see Section 2.3.7) or in suitable wind tunnels, so-called vacuum wind tunnels [2.3.14], [2.3.9]. Due to the low dynamic pressure at high altitudes, vehicle control is only possible using small attitude thrusters because aerodynamic control devices do not operate efficiently in this flow regime. A quick engineering approach to determine the aerodynamic coefficients along a complete reentry trajectory is to apply semi-empirical bridging functions between the rarefied and continuum flow regimes. These bridging functions are derived from computations or measured data. In Figure 2.3.16, an example of this approach is schematically shown for the description of the lift/drag ratio of a vehicle in the flight altitude range of 50–200 km. The bridging function asymptotically approaches the free molecular value (right limit computed by DSMC) as well as the lift/drag ratio determined by the Newton method for the **continuum flow regime** (left limit).

Neglecting viscous effects, the forces acting on a complete configuration are determined by integration of the wall pressure over the entire surface. A characteristic property of hypersonic flight is that in contrast to transonic and low supersonic flight, the lower surface is particularly responsible for the generation of lift. This is due to the high pressures occurring there. Consequently, an inclined flat plate would in practice be the most effective lower surface for a reentry vehicle. By means of the pressure distribution determined by the original Newton method, the following correlations for the lift and drag coefficients for a flat plate result:

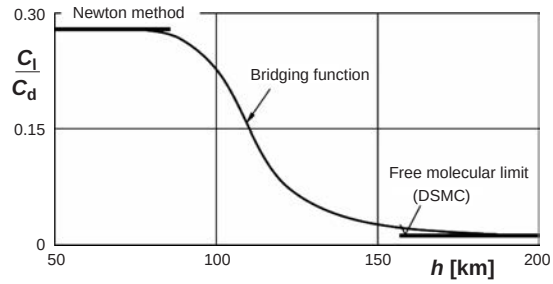


Figure 2.3.16: *Semi-empirical bridging function description of the lift/drag ratio of a vehicle in the flight altitude range of 50–200 km.*

$$C_l = 2 \cdot \sin^2 \alpha \cdot \cos\alpha \qquad (2.3.16)$$

$$C_d = 2 \cdot \sin^3 \alpha \qquad (2.3.17)$$

From the derivative of the correlation for C_l with respect to the angle of attack α, it is found that the maximum lift occurs at an angle of attack of $\alpha_{C_{l,max}} = 54.7°$. In spite of the simplifying assumptions underlying the Newton method, this value represents a realistic estimate which is in the range of angles of attack at which actual lift-generating reentry vehicles reach the maximum lift coefficient.

In contrast to the considerations related to the heat flux into the structure of a reentry vehicle, viscous effects play an inferior role related to the aerodynamic properties of such vehicles. Regarding the pressure drag, fundamental differences between hypersonic and subsonic flight can be identified. In subsonic flight, the pressure drag is significantly determined by the size of the wake, whereas for supersonic and hypersonic vehicles the shape of the nose represents the determining factor. From a drag reduction point of view, vehicles designed for subsonic flight are equipped with tapered trailing edges, and supersonic and hypersonic configurations with pointed leading edges. These considerations demonstrate the basic dependencies and are predominantly valid for hypersonic aircraft designed to achieve a certain flight range. It should be noted that, due to the thermal balance and the thermal limits of currently applied materials, blunt-shaped nose parts are chosen for reentry vehicles. Another aspect is related to the shape of the wing. Straight leading edges are preferred in subsonic flight, because they generate lower induced drag and higher lift. Hypersonic vehicles with high lift/drag ratio, however, exhibit swept-back geometries with sharp leading edges. The delta wing shape with round leading edges, as used on the Space Shuttle Orbiter, represents a compromise which allows sufficient lift to be generated at low speed, to achieve a satisfactory lift/drag ratio at hypersonic speeds and to meet the requirements of thermal balance [2.3.2], [2.3.15]. The position of the aerodynamic center on a delta wing depends on the flight Mach number. While in subsonic flight the aerodynamic center is located at half of the chord of the wing, it is located in the center of the wing area in hypersonic flight. The latter results from the fact that the pressure is almost constant along the lower side of the wing. This discussion emphasizes that the design of a hypersonic aircraft or reentry vehicle which ought to be able to land on a conventional runway must satisfy conflicting requirements regarding optimum aerodynamics in subsonic and hypersonic flight [2.3.12]. The basic differences are schematically summarized in Figure 2.3.17.

2.3.7 Tools for the Determination of Aerothermodynamic Data

The detailed design of hypersonic or reentry configurations is performed by utilizing experimental

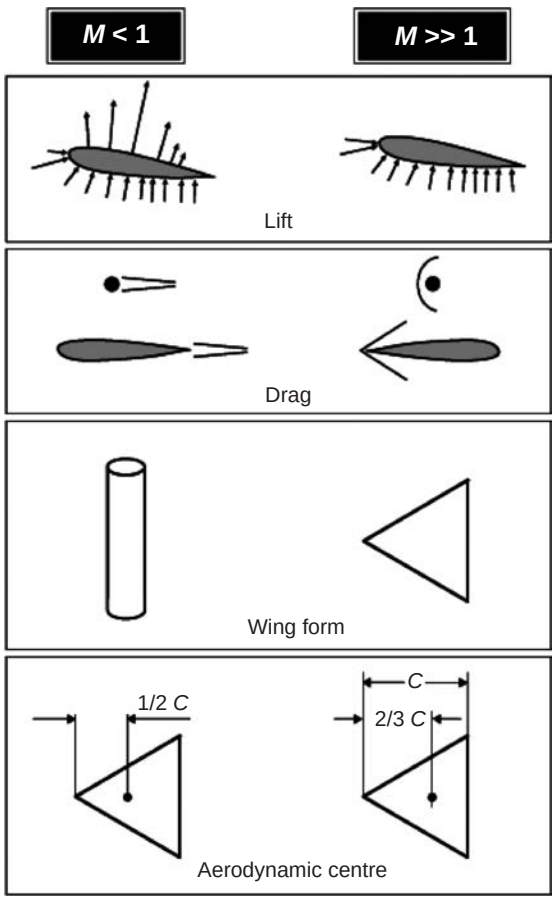


Figure 2.3.17: *Influence of the Mach number on vehicle aerodynamics and vehicle shape [2.3.12].*

data resulting from wind tunnel investigations and flight experiments as well as data from numerical flow field analyses. As flight tests provide the most realistic conditions for the generation of experimental data, they are of great importance for the development process [2.3.16]. However, hypersonic and reentry flight experiments are very complex, generally expensive, and their repetition rate is low. For these reasons, the main sources of aerothermodynamic data are wind tunnel experiments and numerical computations.

The experimental investigation of hypersonic and reentry flows in ground-based tests is performed at a variety of different types of facilities. The reason is the large range of flow conditions and phenomena encountered in hypersonic flight and the fact that no single facility can simulate all relevant flow parameters simultaneously. Therefore, in hypersonic testing, the principle of partial simulation is applied. This implies that different flow phenomena are studied at different types of facilities. Examples are Mach–Reynolds number simulation in cold hypersonic ground-based test facilities, verification and qualification of hot structures of space vehicles in **arc-heated test facilities**, or the investigation of the influence of chemically reacting flows past reentry vehicles on their aerodynamic behavior in **shock tunnels**. Comprehensive overviews of the ground-based testing of hypersonic flows are given in, for example, [2.3.17], [2.3.18] or [2.3.19].

An aspect which exemplifies the partial simulation is that one possibility to increase the Mach number in ground-based facilities is to reduce the free stream temperature, that is the free stream speed of sound. Although in these cold hypersonic testing facilities flight Mach numbers can be reproduced, the free stream velocity in the test section is significantly lower than the actual flight velocity (see Figure 2.3.18 and Figure 2.3.19). However, as pointed out in Section 2.3.3, one characteristic of hypersonic flight at high Mach number is that the kinetic energy of the flow is large enough that, in the shock layer of reentry vehicles, for instance, high-temperature effects such as vibrational excitation or dissociation of the fluid molecules is induced. These effects cannot be duplicated in cold hypersonic ground-based test facilities.

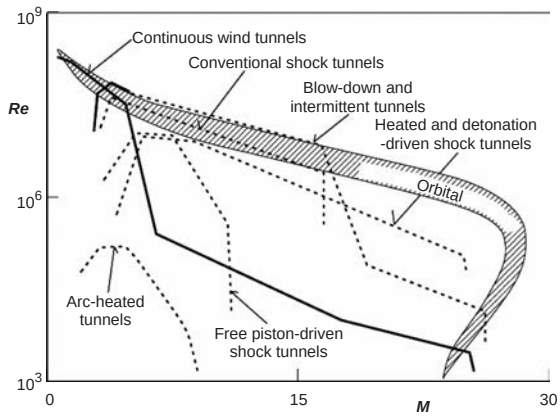


Figure 2.3.18: *Operating range of different types of wind tunnels as functions of Mach and Reynolds numbers; solid and dashed lines represent cold and hot (h_0 < 2 MJ/kg) hypersonic facilities, respectively. The given Reynolds number is based on the radius of the test section.*

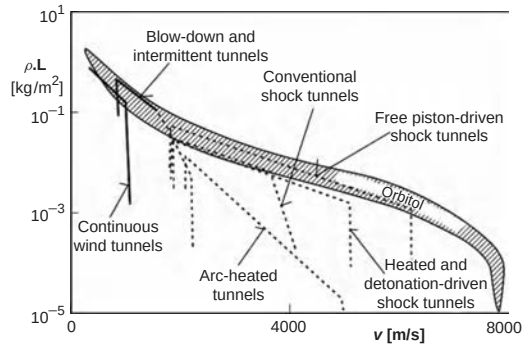


Figure 2.3.19: *Operating range of wind tunnels as a function of the binary scaling parameter and the velocity; solid and dashed lines represent cold and hot (h_0 < 2 MJ/kg) hypersonic facilities, respectively.*

The high energy content of flows during reentry was addressed at the beginning of this section. As can be expected, technical challenges are involved in the experimental reproduction of such flows. The operation of a hypersonic testing facility capable of generating a flow with a free stream velocity of 6000 m/s and a density of 0.003 kg/m^3 in a test section with an area of 1 m^2 would result in a power requirement of 300 MW. Therefore, continuous flow facilities are not a practical way to generate such high enthalpy hypersonic flows. Additionally, correct simulation

in ground-based testing of the chemical relaxation length of the dissociation reactions of the fluid molecules, occurring for example behind the strong bow shock in front of the nose of a reentry vehicle, requires the duplication of the flight binary scaling parameter $\rho \cdot L$. This parameter represents the product of the free stream density ρ_∞ times a characteristic flow length L [2.3.20]. Consequently, the smaller the scale of the wind tunnel model, the higher the free stream density or pressure in the facility test section. The highest local heat flux occurs at a point in the reentry flight trajectory at about 70 km altitude with an associated atmospheric density of approximately 10^{-4} kg/m^3. Scaling the vehicle down by a factor of 30, a free stream density in the ground-based facility of 0.003 kg/m^3 is required. If a flow with this free stream density and a velocity of 6000 m/s is generated by expansion in a convergent–divergent hypersonic nozzle from a reservoir at rest without adding energy, a total specific enthalpy of about 23 MJ/kg and a nozzle reservoir pressure of approximately 90 MPa are required. This results in a nozzle reservoir temperature of about 10 000 K. Hence, ground-based test facilities are required in which high reservoir temperatures and pressures can be generated simultaneously. The operating range of some typical hypersonic facilities related to the reproduction of the flight velocity and the binary scaling parameter is illustrated in Figure 2.3.19. Due to the high thermal loads of wind tunnel components, flows with the conditions described above can only be achieved experimentally in impulse facilities with short flow durations. This relationship is depicted in Figure 2.3.20. In shock tunnels, typical test times in the order of milliseconds are achieved, resulting in specific requirements for the applied measurement techniques. The complexity of high enthalpy flows demands that experimental programs be performed in close coordination with numerical investigations [2.3.21].

Another aspect of experimental aerothermodynamic investigations is the qualification of new materials for **thermal protection systems** or the development and testing of novel cooling concepts under realistic flight conditions. This implies that the surface temperature of wind tunnel models must reach similarly high temperatures – in the order of 2000 K – as obtained in flight. For these investigations, arc-heated

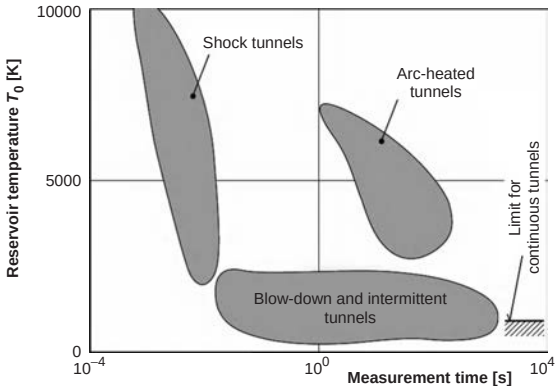


Figure 2.3.20: *Achievable reservoir temperatures and available test times of different types of ground-based hypersonic test facilities.*

facilities are particularly suitable because (as depicted in Figure 2.3.20) their test time is significantly longer than that of shock tunnels and therefore the required surface conditions can be adjusted.

The limits of the different facility-type operating ranges given in Figures 2.3.18 to 2.3.20 are meant to illustrate basic differences. The operating ranges of individual facilities belonging to one of the mentioned groups can differ in detail from the boundaries shown. Likewise, the trajectory of a lift-generating reentry configuration from Low Earth Orbit (LEO) incorporated in Figures 2.3.18 and 2.3.19, simply serves as an orientation. When considering a different type of reentry vehicle or reentry trajectory, it must be taken into account that the relationship to the operating range of the different types of facilities is changed.

Regarding the numerical treatment of aerothermodynamic problems, different approaches are possible. With increasing complexity these include empirical or semi-empirical correlations, surface inclination methods (Newton method etc.) and ultimately complex and detailed computations of three-dimensional steady and unsteady flow fields. An overview of the computational methods commonly used is given in Table 2.3.2. Methods of the first and second levels of complexity are predominantly applied for first qualitative assessments and in the context of preliminary design activities.

During reentry, a spacecraft flies through regimes which are characterized by quite different physical

2

Table 2.3.2: *Comparison of different levels of numerical methods of calculation used in aerothermodynamics.*

Level	Mathematical model	Area of application	Complexity	Required computer time
0	Empirical correlations	Qualitative predictions	Algebraic equations	Seconds
1	Surface inclination methods	Quantitative predictions for pre-design	Algebraic equations	Minutes
2	Euler equation	2D, 3D; continuum flow regime, inviscid flows; no separation	Partial differential equations	Hours
3	Navier–Stokes and Boltzmann equations	2D, 3D; viscous flows, depending on the validation of the models describing turbulence, high-temperature effects and particle collisions	Partial differential equations, stochastic models	Hours up to several days

and chemical phenomena (see also Figure 2.3.1). One example is the transition from free molecular flow to continuum flow described in Section 2.3.3. Figure 2.3.21 illustrates that the **Boltzmann equation** is the only mathematical model capable of covering the complete Knudsen number range relevant for reentry. The direct numerical solution of the Boltzmann equation is very complex, and in general particle methods such as the DSMC method are applied [2.3.22]. DSMC is the only feasible (i.e., economical) numerical method capable of computing flows in the range between the continuum and free molecular regimes, and is thus the method of choice for high-altitude flow field calculations. In the regime of denser atmosphere, application of the Navier–Stokes and Euler equations is more efficient [2.3.23], [2.3.24]. Models based on the **Navier–Stokes equation** have been proved useful up to Knudsen numbers of $Kn = 0.3$ when extended by suitable models such as slip flow. DSMC methods as well as methods for the solution of the

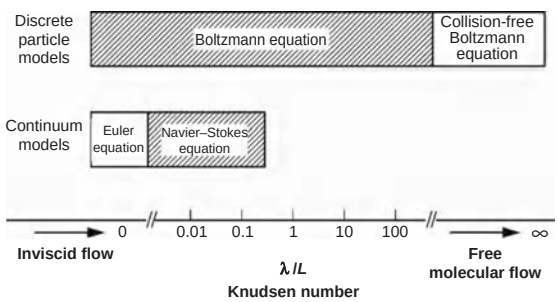


Figure 2.3.21: *Scope of application of different mathematical models as a function of the Knudsen number.*

Navier–Stokes and the **Euler equations** were extended to model high-temperature effects, that is thermal and chemical relaxation processes. In addition to the processes which occur inside the flow field, the correct modeling of wall catalysis is also of particular importance. The direct numerical computation of turbulent, application-oriented flows using Navier–Stokes methods is currently not practical. For the performance of such flow field investigations, the system of equations can be approximately solved by applying turbulence models based on semi-empirical approaches. For the validation of models which describe high-temperature effects or turbulence, appropriate experimental data is necessary. The validation process should in general be performed for each considered class of flow in order to minimize uncertainties related to the predictions resulting from numerical flow field computations.

In summary, detailed aerothermodynamic investigations are only possible if complex numerical methods are utilized so that computations can be conducted on high-performance computers, and if collateral experiments in cold and hot hypersonic ground-based test facilities are carried out. The strategy to extrapolate the experimental data obtained in wind tunnels and the corresponding numerical data to flight conditions must be validated via dedicated flight experiments.

Plotted in Figure 2.3.22 [2.3.4] is the percentage of the three tools to determine aerothermodynamic data relating to their utilization in the configuration design and the subsequent configuration analysis. As an example it should be noted here that for the design of the Space Shuttle Orbiter, 25 000 wind tunnel hours were expended.

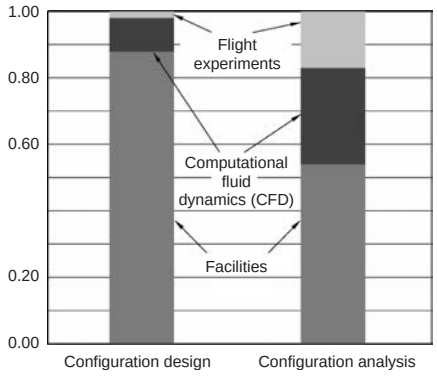


Figure 2.3.22: *Percentage of the three tools to determine aerothermodynamic data related to their utilization in the configuration design and the configuration analysis.*

Bibliography

[2.3.1] Anderson, J.D. (Jr.) *Introduction to Flight*. New York: McGraw-Hill, 1989.

[2.3.2] Anderson, J.D. (Jr.) *Hypersonic and High Temperature Gas Dynamics*. New York: McGraw-Hill International Editions, 1989.

[2.3.3] Isakeit, D., Watillon, P., Wilson, A. *et al.* The Atmospheric Reentry Demonstrator. *European Space Agency Report BR-138*, 1998.

[2.3.4] Longo, J.M.A. Modelling of Hypersonic Flow Phenomena. RTO-EE-AVT 116 *Critical Technologies for Hypersonic Vehicle Development Technology*, AC/323(AVT-116)TP/90, Research and Technology Organization (NATO), 2006.

[2.3.5] Scott, C.D. Wall Catalytic Recombination and Boundary Conditions in Nonequilibrium Hypersonic Flows – With Application. The Third Joint Europe/US Short Course in Hypersonics, RWTH Aachen, 1990.

[2.3.6] Hirschel, E.H. *Basics of Aerothermodynamics*, Progress in Astronautics and Aeronautics. Berlin: Springer Verlag, 2005.

[2.3.7] Brück, S., Kordulla, W., Eggers, Th. *et al.* The Effect of Catalycity on the Heating of the X-38 Shape. Proceedings of the 8th Annual Thermal and Fluids Analysis Workshop on Spacecraft Analysis and Design, University of Clear Lake, USA, 1997.

[2.3.8] Park, C. *Nonequilibrium Hypersonic Aerothermodynamics*. Chichester, John Wiley & Sons, Ltd, 1990.

[2.3.9] Hallmann, W., Ley, W. *Handbuch der Raumfahrttechnik*. Munich: Carl Hanser Verlag, 1988.

[2.3.10] Kolodziej, P. Strategies and Approaches to TPS Design. RTO-EE-AVT 116 *Critical Technologies for Hypersonic Vehicle Development Technology*, AC/323(AVT-116)TP/90, Research and Technology Organization (NATO), 2006.

[2.3.11] Regan, F.J. *Reentry Vehicle Dynamics*, AIAA Education Series. Reston, VA: AIAA, 1984.

[2.3.12] Hankey, W.L. *Reentry Aerodynamics*, AIAA Education Series. Reston, VA: AIAA, 1988.

[2.3.13] Bertin, J.J. *Hypersonic Aerothermodynamics*, AIAA Education Series. Reston, VA: AIAA, 1994.

[2.3.14] Koppenwallner, G., Legge, H. Drag of Bodies in Rarefied Hypersonic Flow. In J.N. Moss and C.D. Scott (eds.), *Thermophysical Aspects of Reentry Flows*, Progress in Astronautics and Aeronautics, Vol. 103. New York: AIAA, 1985, pp. 44–59.

[2.3.15] Whitmore, S.A., Dunbar, B.J. Orbital Space Plane: Past, Present, and Future. AIAA 2003-2718, AIAA International Air and Space Symposium and Exposition: The Next 100 Years, Dayton, OH, July 14–17, 2003.

[2.3.16] Miller, J. *The X-Planes X-1 to X-45*. Hersham, Surrey: Midland Counties Publications, 2001.

[2.3.17] Lukasiewicz, J. *Experimental Methods of Hypersonics*. New York: Marcel Dekker, 1973.

[2.3.18] Lu, F.K., Marren, D.E. (eds.) *Advanced Hypersonic Test Facilities*, Progress in Astronautics and Aeronautics, Vol. 198. New York: AIAA, 2002.

[2.3.19] Stalker, R.J. Modern Development in Hypersonic Wind Tunnels. *Aeronaut. J.*, January, 21–39, 2006.

[2.3.20] Stalker, R.J. Hypervelocity Aerodynamics with Chemical Nonequilibrium. *Annu. Rev. Fluid Mech.*, 21, 37–50, 1989.

[2.3.21] Hannemann, K. High Enthalpy Flows in the HEG Shock Tunnel: Experiment and Numerical Rebuilding. AIAA 2003-0978, 41st Aerospace Sciences Meeting and Exhibit, Reno, ND, January 6–9, 2003.

[2.3.22] Bird, G.A. *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*. Oxford: Oxford University Press, 1994.

[2.3.23] Laney, C.B. *Computational Gas Dynamics*. Cambridge: Cambridge University Press, 1998.

[2.3.24] Toro, E.F. *Riemann Solvers and Numerical Methods for Fluid Dynamics. A Practical Introduction*, Third Edition. Berlin: Springer Verlag, 2006.

2.4 Meteoroids and Space Debris

Hans-Günther Reimerdes

2

Spacecraft are not alone in space during their missions; they share it with meteoroids and man made orbiting objects, called space debris or orbital debris. This poses the possibility of unwanted collisions. Because of the high speed of impact (10 km/s or more) even small particles can cause substantial damage or the loss of a spacecraft. While meteoroids are of natural origin, space debris has been created by the space activities of the last 50 years. The risk posed by space debris is growing continuously and for manned missions in low Earth orbits (LEOs) it is already higher than that resulting from meteoroids. As a consequence, nowadays manned spacecraft have to be protected against impacting objects. In the future this will also be necessary for unmanned missions. Furthermore, future missions have to be planned so that they will not contribute to an uncontrolled increase of space debris.

2.4.1 The Environmental Conditions

2.4.1.1 Meteoroids

Meteoroids result from the disintegration or breakup of comets and asteroids, and they move in orbits around the Sun, passing the Earth. It can be assumed that the total mass in the near-Earth environment (below 2000 km altitude) is about 200 kg in total [2.4.1]. A distinction is made between continuous random meteoroid flux (**sporadic flux**) and flux peaks (**streams**) occurring at regular intervals [2.4.2]. About 10% of the total flux results from the streams. They appear in short time periods and their flux may exceed the sporadic flux by orders of magnitude. The Leonides are a well-known example of streams.

The density of the meteoroid particles varies from 0.15 to 8 g/cm^3. The speed of collision relative to Earth is between 11 and 70 km/s. The average velocity is about 17 km/s [2.4.1].

Based on observations made from Earth and on measurements performed by satellites with dedicated sensors as well as on the evaluation of surfaces from retrieved spacecraft brought back to Earth, models have been developed to describe the space environment. These models are becoming more and more realistic and thus also more complex. The model developed by Grün *et al.* [2.4.3] is suitable for manual computations and first estimations and is used by NASA to describe the meteoroid environment for the International Space Station (ISS) [2.4.1]. The meteoroid module of the ESA environment model MASTER [2.4.4] is more complex. It is based on the Devine model [2.4.5], which was extended by Staubach [2.4.6].

2.4.1.2 Space Debris

While the number of meteoroids remains nearly constant, the number of space debris particles continues to increase with ongoing space mission activities. Many objects remain in space and thus their number increases continuously. Space debris includes objects like upper stages, satellites and payloads, as well as parts released during the missions like bolts, springs or the protective covers of optical instruments. The firing of solid rocket motors (SRMs) produces residues like slag and dust. Fragmentation as a result of explosions or collisions leads to an increased number of particles. Thermal cycling and atomic oxygen corrosion acting on painted surfaces and thermal protective foils lead to the release of chips of coating material. Impacts of small particles create ejection material, and so on. Figure 2.4.1 presents the different orbital debris particles in LEO and their typical sizes. Contrary to the above-mentioned mechanism of space debris creation, there are natural decay processes like atmospheric drag and solar radiation pressure that slowly reduce the number of particles. They change the orbit of the objects until they reenter the Earth's atmosphere [2.4.7]. However, these mechanisms are not able to compensate for the creation of particles, thus in total the number of particles increases.

Most of the data on space debris is collected by the United States Space Command [2.4.8]. It observes objects larger than 10 cm in LEO and larger than 100 cm in geostationary orbit (GEO). The result is a catalog containing about 10 000 objects (as at the middle of 2008 – see also Figure 2.4.2) [2.4.9]. Of these objects about 3000 are spacecraft, 1600 are stages

from launchers and 5400 are fragments, mainly caused by explosions. Figure 2.4.2 shows the continuous increase in the number of fragments. It can be noted that the rate of object decay increases within the 11-year solar cycle at times of increased solar activity. In addition to the observed objects there are about 110 000 objects between 1 and 10 cm in size, and the number of objects smaller than 1 cm is estimated to be about 35 million [2.4.10]. Figure 2.4.3 shows the number of particles per cubic kilometer as a function of altitude for three different particle sizes (computed with MASTER 2005 [2.4.4]). There are three regions of particle concentrations, representing the main used orbits: the LEO below 2000 km altitude, the 12 h orbits at about 18 000 km and the GEO (24 h orbit) at about 36 000 km altitude. In LEO the amount of space debris is much larger than that of meteoroids, except for particle sizes below 0.1 mm where the meteoroids are predominant.

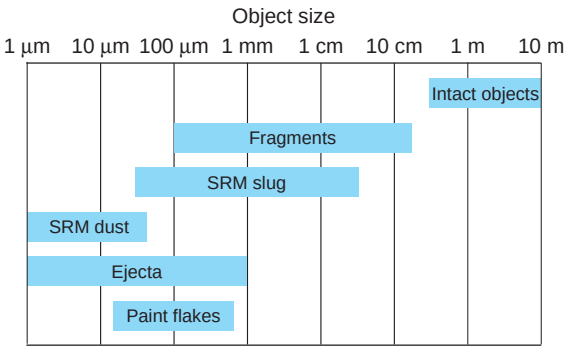


Figure 2.4.1: *Size of space debris particles in LEOs.*

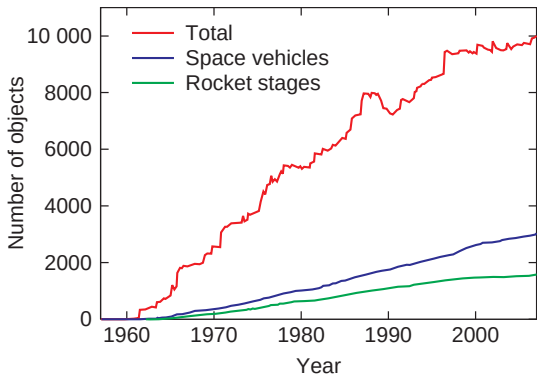


Figure 2.4.2: *Regularly observed objects [2.4.9].*

The density of space debris corresponds to that of the materials used for spacecraft. It is between 1.8 and 8.9 g/cm^3. About 50 to 60% of the particles have a density of about 2.8 g/cm^3 (aluminum alloys and glass) [2.4.1]. Their velocity with respect to the Earth is around 7–8 km/s in LEO, resulting in collision velocities of up to about 16 km/s.

In recent years models have been developed to describe the **space debris environment**. For planned missions they allow computation of the risk of being hit by particles. Frequently used models are the ESA MASTER model and the ORDEM (Orbital Debris Environmental Model) of NASA [2.4.11]. These models are regularly updated and become more realistic with time, but also more complex. Nowadays, special computer software (provided by space agencies) is needed to compute the environmental data for a given mission. Again, the models are based on radar measurements made from Earth, on measurements made on-board a spacecraft, as well as on the evaluation of surfaces of retrieved spacecraft brought back to Earth.

2.4.2 Future Development and Debris Mitigation Measures

As mentioned in Section 2.4.1.2, the amount of space debris increases and thus so does the risk for future missions. Experts forecast a scenario [2.4.12] where

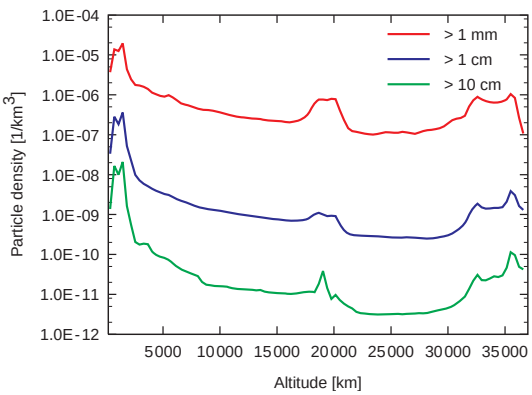


Figure 2.4.3: *Space debris density as a function of altitude (MASTER 2005).*

2

the spatial density of objects achieves a critical value. This is the case when collisions between objects result in a chain reaction and an exponential growth of fragments. To avoid this, the space agencies are developing mitigation measures in international committees and providing recommendations [2.4.13] for the planning of future missions.

These considerations are based on predictions for the development of the space debris environment in the next 50 to 100 years. The first main contributions were given by Rex [2.4.7], who conducted basic research in this field in the early 1980s. This early research was also the basis for the environment model MASTER mentioned above. Investigations performed for and by ESA [2.4.10] predict an increase of objects larger than 10 cm by a factor of 2.5 within the next 100 years (Figure 2.4.4). This prediction is based on the assumption that space activities will continue in the future as at present (business as usual). For particles larger than 1 cm, an increase by a factor of 5 is predicted for the same time period (Figure 2.4.5). Here mainly collisions contribute to this increase.

According to Rex [2.4.7], uncontrolled growth in the amount of space debris will make safe operation in space impossible in the future. This prospect leads to considerations of how to avoid collisions. Special missions to **collect space debris particles** in order to clean up the environment cannot be realized from a cost point of view. Two other measures are presently considered to have a reasonable chance of realization [2.4.10]:

- Avoidance of explosions in orbits by passivation of inactive stages and satellites
- Deorbiting objects at the end of their lifetime.

Up to now, explosions have mainly contributed to the generation of space debris. To avoid these in future, the remaining fuel should be released from inactive upper stages and satellites. This is called **passivation** and is already being done. However, if the passivated objects remain in space, the possibility of collision is still present. To improve this situation, stages and satellites should be deorbited at the end of their mission by distinct "delta-v" maneuvers. This would be an effective way to keep space clean for future space activities [2.4.7]. For orbits with increasing altitudes this method becomes very expensive and cannot be easily realized. A compromise is to reduce the altitude by a single "delta-v" maneuver and then make use of the natural drag of the atmosphere to slowly decrease the altitude.

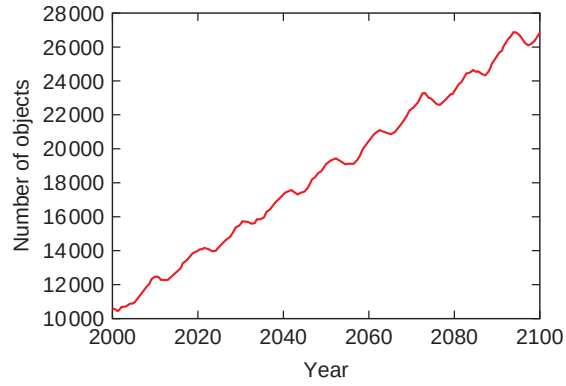


Figure 2.4.4: *Prediction of the development of objects larger than 10 cm [2.4.10].*

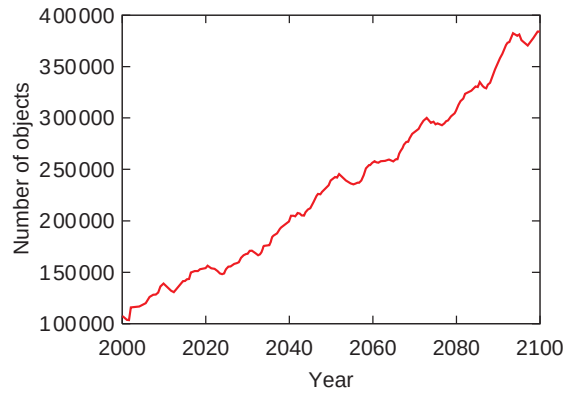


Figure 2.4.5: *Prediction of the development of objects larger than 1 cm [2.4.10]*

Computations performed at ESA/ESOC [2.4.10] show the expected effect of passivation (Figure 2.4.6) and deorbiting (Figure 2.4.7) on the development of the number of particles within the next 100 years. From these investigations a recommendation is derived to always perform passivation in the future and to plan missions such that deorbiting takes place within 25 years. Deorbiting can only be implemented for missions in LEO. For orbits at higher altitudes, for example the 24 h orbits, inactive satellites are placed into so-called **graveyard orbits**, which are 300–400 km above the used orbits.

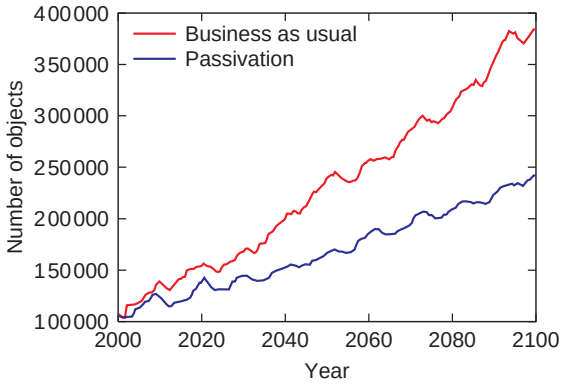


Figure 2.4.6: *Prediction of the development of objects larger than 1 cm for the passivation scenario [2.4.10].*

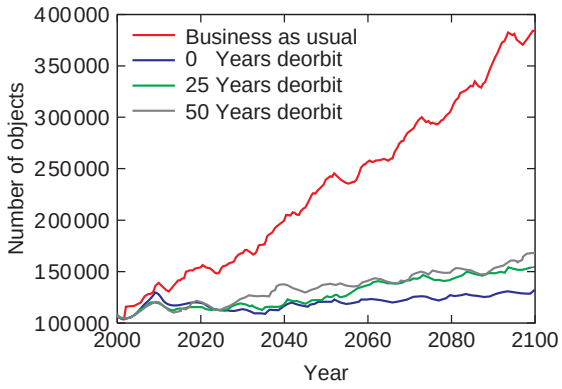


Figure 2.4.7: *Prediction of the development of objects larger than 1 cm for the deorbiting scenario [2.4.10].*

2.4.3 Impact Flux and Impact Risk

An analysis of the risk resulting from meteoroids and space debris is based on environmental models. These models must not only describe the present situation, but also predict future developments. The first environmental models were quite simple [2.4.2], [2.4.14], but with increasing knowledge they became more complex. This is especially the case for space debris, as the continuous use of space results in a permanent change in the environmental conditions. The predictions of the different developed models are compared by international committees (e.g., the Inter-Agency Space Debris Coordination Committee (IADC)) in order to arrive at a common understanding.

The main value needed for computation of the collision risk is the accumulated flux. It is dependent on the particle size and is defined as the number of particles per area and time. Accumulated flux means all particles larger than a certain size (e.g., with diameter d). With the flux F known, the expected number of impacts N is computed by multiplication of the flux by the duration of the mission T and the area of the spacecraft A:

$$N = FA\,T \tag{2.4.1}$$

The probability of n impacts is computed based on Poisson statistics:

$$p_n = \frac{N^n}{n!}\,\mathrm{e}^{-N} \tag{2.4.2}$$

This gives, for the probability of no penetrations,

$$p_0 = \mathrm{e}^{-N} \tag{2.4.3}$$

The probability of impacts is therefore

$$p_i = 1 - p_0 = 1 - \mathrm{e}^{-N} \tag{2.4.4}$$

If N is very small compared to one, the following approximations hold:

$$\mathrm{e}^{-N} = 1 - N \tag{2.4.5}$$

$$p_0 = 1 - N \tag{2.4.6}$$

$$p_i = N = FAT \tag{2.4.7}$$

The flux may be related to the cross-sectional area (A_c) of the spacecraft, the cross-sectional flux F_c, or to the surface area (A_s), the surface area flux F_s. For convex spacecraft geometries, it can be stated [2.4.1] that

$$F_c = 4\,F_s \tag{2.4.8}$$

If the geometry of the spacecraft and the duration of the mission are known, one can compute for a required probability of no penetration or no failure p_0 the allowed flux which just barely fulfills the requirement:

$$F_s = \frac{(1 - p_0)}{A_s\,T} \tag{2.4.9}$$

If the flux is known as a function of particle size (environment model), the particle size which the spacecraft

has to be protected against can be computed. This computation does not consider that different surface areas of the spacecraft are exposed to different particle fluxes. The flux to a part of the surface A_{s_i} may be a multiple k_i (flux concentration factor) of the average flux. This gives the following number of impacts for a spacecraft assembled from i surfaces:

$$N = F_s\, T \sum_i k_i\, A_{s_i} \tag{2.4.10}$$

The flux concentration factor k_i depends on the orientation of the surface A_{s_i}, on the flight direction and on the directional distribution of the flux. To allow analysis of the **impact risk**, the environment models have to provide the following for an orbit with altitude H, eccentricity e and inclination i:

- Accumulated flux as a function of particle size
- Particle velocity distribution
- Directional distribution.

2.4.3.1 Meteoroids

The risk resulting from meteoroids is substantially smaller than that from space debris. For this reason the relatively simple model by Grün *et al.* [2.4.3] is introduced in the context of this handbook.

It is assumed that the sporadic flux is omnidirectional relative to the Earth. Due to the gravitational field of the Earth, the deep space flux (F_{ip}) increases towards the Earth. This is taken into account with a **focusing factor** $G_\oplus$:

$$G_\oplus = 1 + \frac{R_\oplus}{r} \tag{2.4.11}$$

with:

$R_\oplus$ = Earth radius + 100 km (atmosphere),
r = orbit radius.

A spacecraft in an orbit around the Earth is partially shielded from impacts by meteoroids. This **shielding** increases with decreasing altitude of the orbit and is described by the shielding factor s_f:

$$s_\mathrm{f} = \frac{(1 + \cos\eta)}{2} \tag{2.4.12}$$

$$\text{with } \sin\eta = \frac{R_\oplus}{r}\,.$$

This results in the following flux depending on the altitude H of the orbit:

$$F_\mathrm{h} = s_\mathrm{f}\, G_\oplus\, F_\mathrm{ip} \tag{2.4.13}$$

The velocity of meteoroids relative to the Earth is between 11 and 72 km/s with an average value of 17 km/s [2.4.1]. The model gives the flux as function of the particle mass. With an assumed density of 1 g/cm^3 and spherical particles the interplanetary surface area flux F_{sip} is calculated as in Figure 2.4.8.

Although it is assumed that the meteoroid flux is omnidirectional relative to the Earth, it becomes directional relative to a spacecraft moving through the environment. This is shown in Figure 2.4.9. Here the flux concentration factor k is presented over the surface of a spherical spacecraft at 400 km altitude in a circular orbit with fixed orientation relative to the Earth. Due to the movement of the vehicle in the particle field, a concentration is observed at those surfaces facing the flight direction. The surfaces facing the rear experience substantially smaller fluxes. Due to the Earth's shielding, the fluxes are also very small on the lower surfaces.

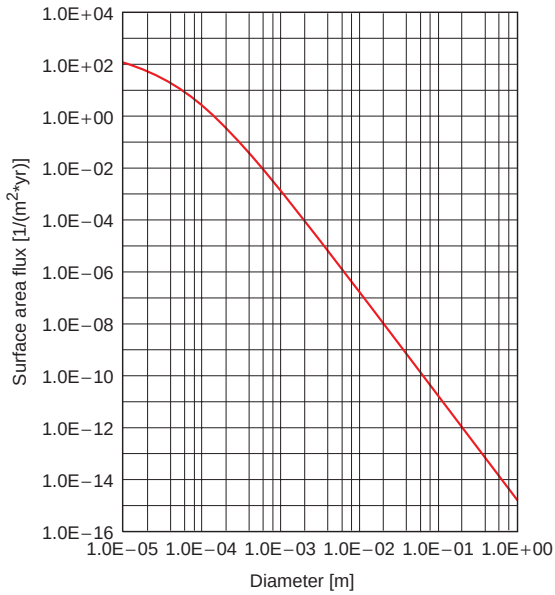


Figure 2.4.8: *Accumulated interplanetary surface area flux F_{sip} as a function of particle diameter of meteoroids [2.4.3].*

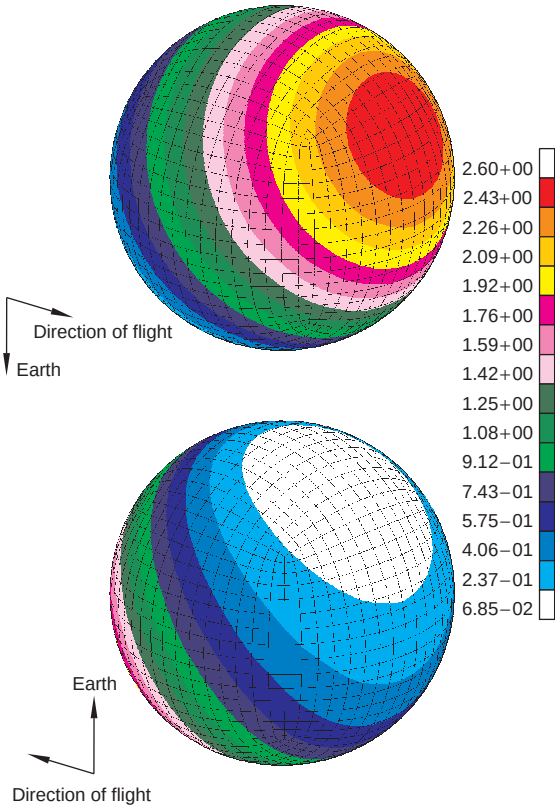


Figure 2.4.9: *Flux concentration factor k as a function of the surface of a spacecraft in the meteoroid environment (H = 400 km).*

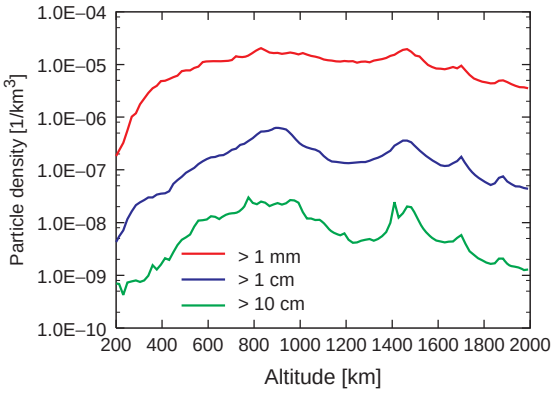


Figure 2.4.10: *Space debris spatial density as a function of altitude (MASTER 2005).*

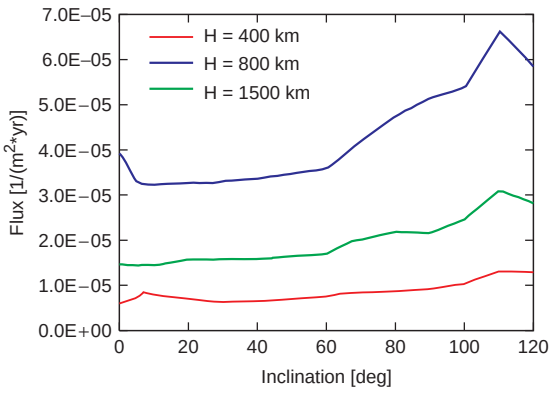


Figure 2.4.11: *Space debris flux as a function of inclination (d > 1 cm) [2.4.10].*

2.4.3.2 Space Debris

Space debris is substantially more dangerous for spacecraft in near-Earth orbits than are meteoroids. As a result of space activities in frequently used orbits (height, inclination) the particle flux exhibits pronounced directions and concentrations. These must be described by the models, which lead to very complex computations possible only with dedicated software. Here some of the most significant characteristics of the space debris environment are described. The data presented was computed using the ESA's MASTER 2005 [2.4.4] and NASA's ORDEM2000 [2.4.11].

Figure 2.4.10 shows **spatial densities of particles** larger than 1 mm, larger than 1 cm and larger than 10 cm from 200 to 2000 km altitude. There are concentrations of particles at approx. 800 km and approx. 1500 km altitude. The number of particles there is about 10 times the number at 400 km altitude, the orbit of the ISS. The large number of missions in polar orbits led to larger fluxes being observed between inclinations of 80° to 120° than between 20° and 60° (Figure 2.4.11). It follows that in orbits with increasing inclination the average impact velocity increases (Figure 2.4.12). This amounts to approx. 10 to 12 km/s. The maximum impact velocity is, assuming circular orbits, twice the spacecraft velocity, which amounts to about 7.7 km/s at 400 km altitude.

For the orbit of the ISS with $H = 400$ km and $i = 51.6°$ the surface area flux as a function of particle diameter as computed with ORDEM2000 for the year 2010 is presented in Figure 2.4.13. For comparison the meteoroid flux is also shown.

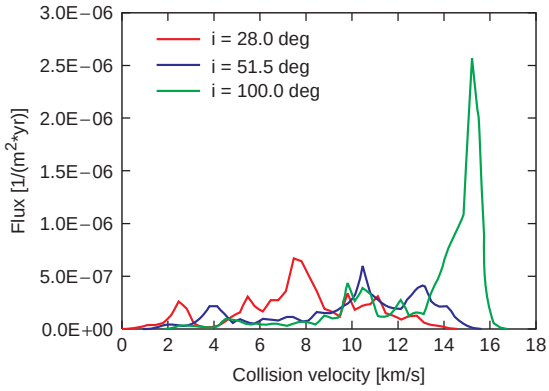


Figure 2.4.12: *Space debris flux as a function of impact velocity for different inclinations (d > 1 cm) [2.4.10].*

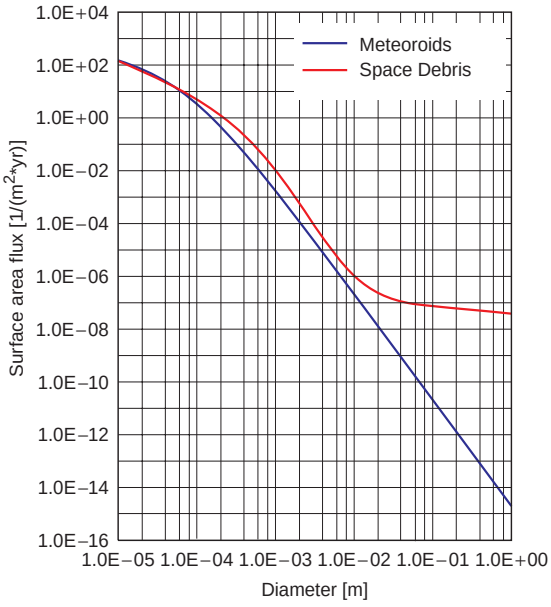


Figure 2.4.13: *Accumulated surface area flux as a function of particle diameter in the year 2010 (H = 400 km, i = 51.6°, ORDEM2000 [2.4.11]).*

Figure 2.4.14 shows the **flux concentration factor** for a spherical spacecraft in the space debris environment. The result is based on ORDEM2000 for particles with a diameter larger than 1 cm. The highest flux concentration again results for surfaces in the flight direction. Since the majority of particles are in circular or elliptical orbits with small eccentricity,

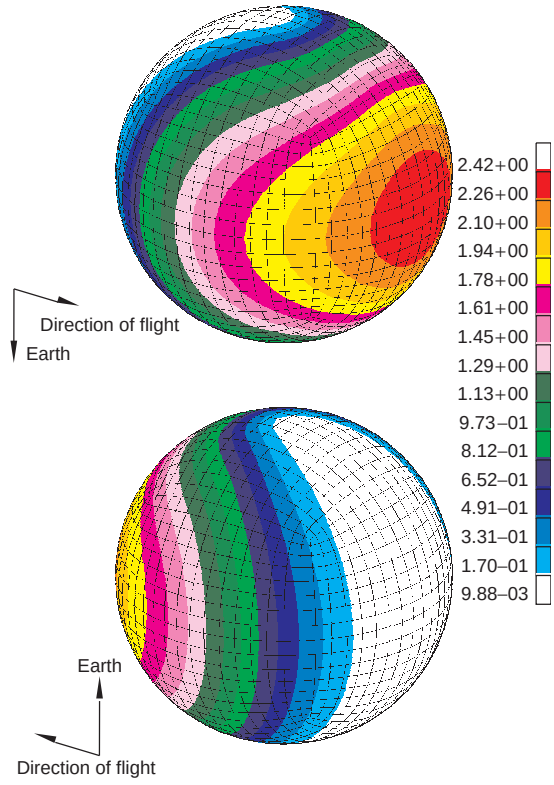


Figure 2.4.14: *Flux concentration factor k as a function of the surface of a spacecraft in the space debris environment (H = 400 km, i = 51.6°).*

the surfaces parallel to the Earth's surface and on the rear experience only very few impacts.

2.4.3.3 Impact Risk

From Figure 2.4.13 it can be seen that the impact risk due to space debris is larger than that due to meteoroids if the particles are larger than 0.1 mm. With the well-known flux as a function of diameter it is possible to assess in which time intervals **impacts of particles of a certain size** are to be expected, using Equation 2.4.7. Assuming a spacecraft with a surface area of 150 m^2 in a near-Earth orbit with $H = 400$ km and $i = 51.6°$, the time intervals given in Table 2.4.1 are computed for different particle sizes. These time intervals represent statistically determined values. Even if the statistically determined time interval is very large (e.g., for particles larger than 1 cm), a collision

Table 2.4.1: *Time interval between impacts of objects with diameter d.*

D	> 0.1 mm	> 1 mm	> 1 cm
Meteoroids	0.68 days	3.43 years	34 347 years
Space debris	0.49 days	0.53 years	4 444 years

may happen after a much shorter time. This was the case in the collision of the satellite CERISE relatively shortly after its operational start with a cataloged fragment of an Ariane upper stage [2.4.15].

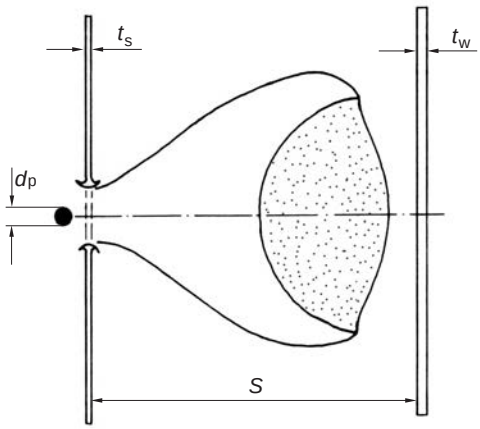


Figure 2.4.15: *Effect of a protection shield.*

2.4.4 Protection of Spacecraft Against Impacting Particles

Particles which impact with speeds of 10 km/s or higher on metallic structures produce a crater with a depth of about 3–5 times the particle diameter [2.4.16]. A thin metal wall will be perforated if the crater depth is about 60% of the wall thickness. A 1 mm particle is able to perforate a wall of 5 mm thickness. To withstand the mechanical loads, the structure of a spacecraft has wall thicknesses of 1 to 2 mm. From this it becomes evident that protective measures are necessary if a certain impact risk exists.

2.4.4.1 Protection Concepts

The simplest protection concept is to increase the wall thickness of the outside structure of the spacecraft. However, that leads very quickly to an unacceptable increase in structural mass. The introduction of a **protective shield** in front of the structure is substantially more effective. This was suggested in 1947 by Whipple [2.4.17], which led to the **Whipple shield** or the bumper shield. Its efficiency results from the fact that an incoming particle is destroyed upon impact. The fragments heat up and may melt or even vaporize [2.4.18], [2.4.19]. The cloud of fragments, droplets or vapor produced between the two walls expands and hits the second wall (protected structure) over an enlarged area (Figure 2.4.15). The impact energy is distributed over a larger surface and is therefore less damaging. In particular, protection concepts with one or more shields have been developed for the manned modules of the ISS in the last 15 years [2.4.20], [2.4.21], [2.4.22], [2.4.23]. These provide protection against particles with diameters up to 1 cm. In addition to the outer shield these advanced concepts have an internal shield as well.

The concepts described above are particularly intended for the protection of manned spacecraft in order to ensure sufficient security for the crew. For unmanned spacecraft with mostly smaller surfaces and less stringent safety requirements, it is sufficient to plan for protection against particles of 2–3 mm diameter. The additional costs for protection should be as low as possible. A meaningful solution is to design the load-carrying structure as a **sandwich structure**. This not only provides good structural characteristics and low mass, but also offers better protection than monolithic structures. Compared to a **double wall** the sandwich is, however, less efficient, since the core reduces the expansion of the particle cloud between the face sheets. Another option is **multilayer insulation** (MLI), which is introduced for passive temperature control in front of the structure and behaves like a very thin protective shield. In the context of a technology study performed on behalf of ESA/ESTEC [2.4.24], economical protection concepts for satellites have been developed; these are reinforced MLI and sandwiches with enhanced protective capability.

2

2.4.4.2 Design of Protection Concepts

For the design of protective measures both the meteoroid and space debris environments must be known, as well as the behavior of the protection concepts. This is described by **damage or ballistic limits equations**. With the environment models, the geometry of the spacecraft, the damage equations and the mission data, the protective measures can be evaluated in the context of damage prediction.

Damage Equations

Damage equations describe the behavior of protection concepts, of structures or of components under impact loads. They consider the size, the mass, the velocity and the direction of the impacting particle as well as the geometry and the material of the protection concept. Because of the complex physical phenomena arising during **hypervelocity impacts**, the damage equations are based primarily on test results. One problem is that the velocities attainable in experiments ($v < 9$ km/s) are smaller than those in reality. Extrapolations beyond the realizable velocities are made either with the help of simplified **physical models** or by means of **numeric simulations**. The data achieved by experiments is not sufficient by far to develop general damage equations for all protection concepts. A summary of existing damage equations is given in the IADC Protection Manual [2.4.25]. Experimental techniques and their limits are also documented there. In this handbook damage equations for single-wall and double-wall structures are presented. Equations for more complex configurations are often developed on the basis of these equations. The damage equations are valid for spherical particles.

Single-Wall Structures

The behavior of single-wall structures during hypervelocity impacts depends on the strength and toughness of the impacted structure. One differentiates between **ductile and brittle materials**. The metals used in space structures are ductile, while glass and fiber-reinforced plastics show brittle behavior.

Metals: If a particle hits a very thick metal wall, high pressures and high temperatures develop, fragmenting and melting the wall as well as the particle material. A crater develops with its volume being a multiple of the particle volume (Figure 2.4.16).

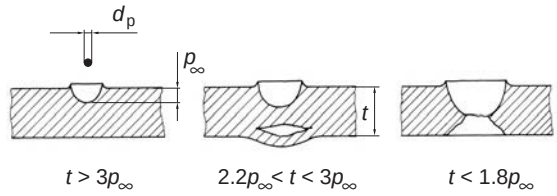


Figure 2.4.16: *Impact on single-wall structures (thin-sheet impact).*

Material is ejected opposite to the direction of impact (ejecta). The high pressure travels as a shock wave through the material, is reflected at the rear side of the wall and returns as a tension wave. If the impacted wall is thin, the tension wave exceeds the tensile strength of the material, leading to **spallation**. **Detached spall** may also threaten components behind the wall.

The **crater depth** p_∞ (in cm) in very thick material is, according to Cour-Palais [2.4.26],

$$p_\infty = 5.24\, H^{-1/4} \left(\frac{\rho_p}{\rho_t}\right)^{\alpha} \left(\frac{v_n}{C}\right)^{2/3} d^{19/18} \qquad (2.4.14)$$

with:

d = particle diameter [cm],
H = Brinell hardness of the wall,
ρ_p = density of the particle [g/cm^3],
ρ_t = density of the wall [g/cm^3],
v_n = impact velocity normal to the surface [km/s],
C = speed of sound of the wall material [km/s].

The exponent α is 1/2 for $\rho_p/\rho_t < 1.5$ and 2/3 for $\rho_p/\rho_t \geq 1.5$. If the thickness of the impacted structure is less than three times the crater depth p_∞, the damages presented in Figure 2.4.16 result. A thin wall will not quite be perforated if the wall thickness t is

$$t = 1.8\, p_\infty \qquad (2.4.15)$$

Figure 2.4.17 shows the thickness required of an aluminum wall (Al 2024–T3) to avoid perforation at an impact velocity of 10 km/s as a function of the diameter of an aluminum particle (space debris).

Glass: Brittle materials with low tensile strength are very sensitive to impact. This characteristic leads to regular replacement of the windows of the Space Shuttle [2.4.27]. In contrast to metals, a flat crater

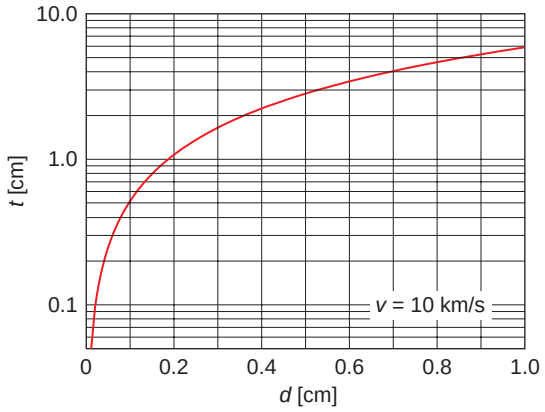


Figure 2.4.17: *Single-wall structure: wall thickness required for protection against space debris impact.*

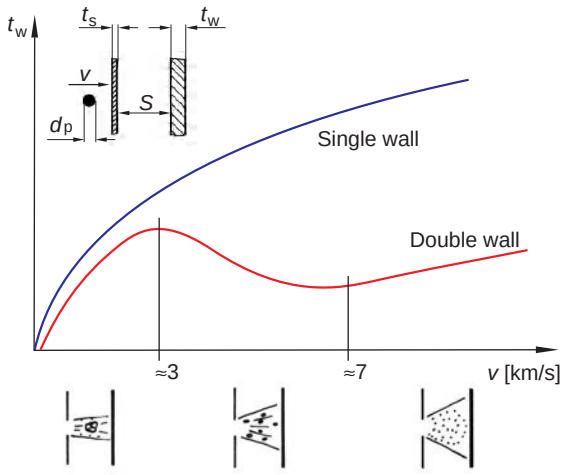


Figure 2.4.18: *Double-wall structure: the effect of impact velocity.*

is created while more material is ejected opposite to the impact direction [2.4.26]. Moreover, a very large wall thickness is necessary, in order to avoid spallation at the rear surface. According to [2.4.26] the crater depth in glass is

$$p = 0.53\ \rho_p^{1/2}\ v_n^{2/3}\ d^{1.06} \qquad (2.4.16)$$

the thickness required to prevent cracks is

$$t_c = 0.14\ p\ v_n^{1.28} \qquad (2.4.17)$$

and to avoid spallation at the rear surface, the thickness should be

$$t_s = 7\ p \qquad (2.4.18)$$

Fiber-reinforced plastics: Fiber-reinforced plastics also show brittle behavior. Here delaminations of the single layers occur, in particular at the rear surface. Their dimensions are substantially larger than those of the front crater [2.4.28]. Generally accepted damage equations for this material are not yet available.

Double-Wall Structures

The geometry of a double-wall structure is depicted in Figure 2.4.18. A shield is arranged at a distance S (spacing) in front of the protected structure (backup wall). The effectiveness of this protection concept depends on the impact velocity. The velocity range is divided into three regions with different behaviors (Figure 2.4.18). At low velocities the pressure developing when the shield is perforated is not sufficient to destroy the particle. There is practically no protective effect. Starting from a critical speed, which depends on the material combination of the shield and the particle, the particle is destroyed. This process increases with increasing velocity. The necessary rear wall thickness decreases at the same time. Above a second critical velocity the particle and its fragments are heated to such an extent that they begin to melt or evaporate. The cloud now represents a distributed load per unit area for the rear wall, its impulse increasing with the impact velocity. According to Christiansen [2.4.20], the critical velocities for shields and particles made of aluminum alloy are v_{lim1} = 3 km/s and v_{lim2} = 7 km/s. The damage equations are given for the velocity from 0 to v_{lim1} and for velocities exceeding v_{lim2}. In the velocity range between the two critical values a linear interpolation is used (see Figure 2.4.19).

For double-wall structures made of aluminum alloys the damage equations proposed by Christiansen [2.4.20] and quoted below are regarded as standard. The **critical particle diameter** d_p (in cm) which can be defeated by the given shielding without detaching spall at the rear wall is computed.

For $v_n < 3$ km/s:

$$d_p = \left(\frac{t_w \left(\frac{\sigma_w}{40} \right)^{1/2} + t_s}{0.6\,(\cos\alpha)^{5/3}\,\rho_p{}^{1/2}\,v^{2/3}} \right)^{18/19} \qquad (2.4.19)$$

with:

t_s = shield thickness [cm],
t_w = backup wall thickness [cm],
S = spacing [cm],
α = impact angle,
v = impact velocity [km/s]
σ_w = yield strength of the backup wall [ksi].

For $v_n > 7$ km/s:

$$d_p = 3.918\; t_w^{2/3}\; \rho_p^{-1/3}\; \rho_s^{-1/9}\; v_n^{-2/3}\; S^{1/3} \left(\frac{\sigma_w}{70} \right)^{1/3} \qquad (2.4.20)$$

The equations are valid if the following equation for the shield thickness is fulfilled:

$$t_s \geq 0.25\; d_p\; \frac{\rho_p}{\rho_s} \qquad (2.4.21)$$

For a double-wall structure with a 2 mm shield thickness, 100 mm spacing and 4 mm backup wall thickness, the critical particle diameter (space debris) is computed by Equations 2.4.19 and 2.4.20 and presented in Figure 2.4.19. The critical particle diameter for a single-wall structure (4 mm wall thickness) is also presented in Figure 2.4.19. It shows the advantage of a protective shield for velocities above 3 km/s.

If a protection concept has to be designed to protect against a given particle, Equations 2.4.19 and 2.4.20 can be used in order to compute the needed wall thickness. In Figure 2.4.20 the wall thicknesses are shown as a function of the particle diameter for an impact velocity of 10 km/s. The distance between the walls amounts to 20 times the particle diameter. The structure consists of aluminum alloy Al 7075 and the density of the particle is 2.8 g/cm^3 (space debris).

Open Questions

The equations given above form the basis for the development of new damage equations. In [2.4.29] a modification is given that allows to describe the influence of the shield thickness on the protective effect. Thus a **mass-efficient shield arrangement** can be designed by means of mathematical optimization [2.4.30]. Damage equations for complex protective systems with more than one shield [2.4.20] or for sandwich structures [2.4.31] are similarly developed. Here the critical velocities, the factors and the exponents are adapted to the results of experiments performed with these configurations. This happens typically during the development of projects for which protection concepts are to be designed. That leads to equations which are not generally valid and applicable only to similar configurations.

The damage equations used today are valid for spherical particles. Recent work shows that the protective effect of a shield is smaller with particles

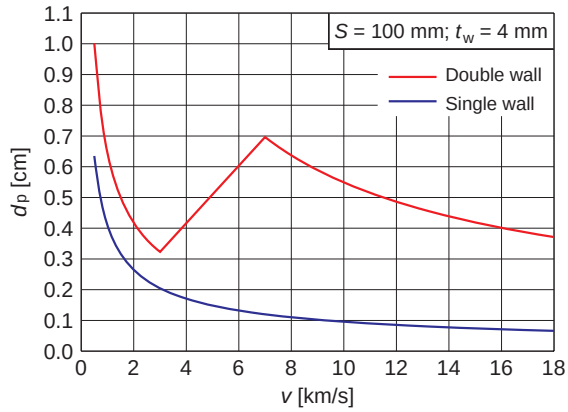


Figure 2.4.19: *Critical space debris particle diameter.*

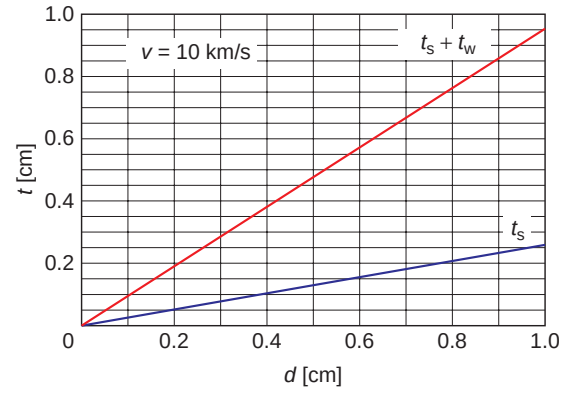


Figure 2.4.20: *Double-wall structure: wall thicknesses required for protection against space debris.*

which are not spheres [2.4.32]. It is expected that future damage equations will consider **particle shape** as well. Which particle shapes are to be expected has to be described by the environment models.

A further uncertainty in the damage equations used is the behavior at **impact velocities** above today's existing test options ($v > 8$ km/s). Although higher velocities can be simulated numerically [2.4.33], the result of these computations depends on the correct description of the material behavior. This can only be determined reliably by experiments, which are again limited to velocities up to 8 km/s.

Damage Prediction

A spacecraft in the meteoroid/space debris environment is to be designed such that during the mission lifetime a required **probability of survival** is reached. This is the probability of no impact leading to failure. The calculation of the protection concept is done in two steps:

- Preliminary design
- Numerical computation using dedicated software for damage prediction.

In both cases environment models and damage equations are needed.

Preliminary Design

In a first step the particle flux is computed, leading to the required probability of no failure (e.g., no penetration, see Equation 2.4.9). The mission duration and the surface area of the spacecraft have to be known. With the flux known the particle diameter leading to this flux can be derived from the environment models (Figure 2.4.13). This is the **particle size** for which the protection concept has to be designed. If the structure is known its damage equations are used to compute whether the particle leads to a failure or not considering the average impact velocity. In case of a failure the structure has to be modified accordingly. If the attitude of the spacecraft is fixed with respect to Earth, then the flux varies over the surface (Figures 2.4.9 and 2.4.14). It may be sufficient to protect only those surfaces with high flux concentration factors k.

The protection concept will be efficient when the number of impacts divided by the surface area is equally distributed over the spacecraft. For a patch i on the surface, the following flux is computed, if the flux concentration factor k_i is known:

$$F_{S_i} = \frac{1 - p_0}{k_i A T} \tag{2.4.22}$$

Thus the particle diameter d_i can be determined against which the protection concept of the patch i is to be designed.

Numerical Damage Prediction

The presented preliminary design does not consider the velocity distribution and the impact angle of the particles. Moreover, the orientation of the surface patches is not considered exactly. In the case of complex geometries, parts of the surface area may be shaded against impact. These influences are considered by **computer programs for damage prediction**. Here the geometry of the spacecraft and its attitude in space are described by surface elements. Damage equations are assigned to surfaces depending on their wall design. From the environment models it can be determined which particles from which directions and with which velocities hit the spacecraft or the individual surfaces. The result of the computation is the number of impacts leading to a failure for the individual surfaces and the probability of survival of the spacecraft (Figure 2.4.21). If the intended protection concept does not lead to a sufficient probability of

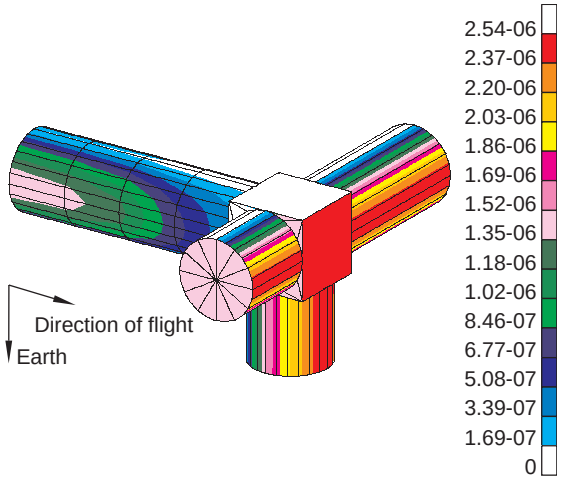


Figure 2.4.21: *Expected number of impacts on a spacecraft per unit area (space debris, dp > 1 cm).*

survival, the results allow identification of where an improvement to the protection concept should be made. In connection with mathematical optimization and suitable damage equations [2.4.30], the protection concept can be optimized in such a way that the required probability of survival is obtained with a minimum of additional mass.

2.4.5 Mission Planning

With respect to meteoroids and space debris, two substantial aspects are to be considered when planning a mission. These are on the one hand avoiding the generation of new space debris and on the other hand reducing the risk resulting from impacts.

Avoiding Space Debris: In order to avoid increasing the amount of space debris, the following aspects are to be considered during mission planning. It has to be guaranteed that the new spacecraft does not remain in orbit for longer than 25 years. Depending on the orbital altitude, this is achieved either by deorbiting the spacecraft or by reorbiting it into a graveyard orbit. The chosen launcher should fulfill the standards of avoidance. This means passivation of upper stages and their deorbiting back to Earth. Solid propellant motors should be designed to release no or only little dust and slag. Mission-related objects have to be prevented from becoming debris. This means that cover caps and pyrotechnic devices should be attached to the spacecraft so that they are not released into space after use. Ductile materials are to be used for surface coatings in order to minimize blistering and the generation of paint flakes.

Reducing the Risk: In order to minimize the risk of particle impacts, the spacecraft is to be built as small and compact as possible. If the planned mission permits, the height and the inclination of the orbit should be selected so that the particle fluxes to be expected are as small as possible. Short-duration missions (e.g., Space Shuttle missions) should be planned for times when no meteoroid streams are expected. If the vehicle's orientation is fixed with respect to the Earth's surface an attitude is favorable where surfaces exposed to high fluxes are as small as possible. Safety can be increased substantially by the use of multiwall instead of single-wall structures. A sandwich structure is preferred to a single-wall structure. With its face sheets it practically represents a two-wall structure. Sensitive and critical components like payloads, receivers, batteries and pipes should be placed on the inside far away from critical surfaces (with high fluxes). If the architecture of the spacecraft does not permit this, these components may need to be additionally protected.

Bibliography

[2.4.1] Anonymous. Space Station Program Natural Environment Definition for Design. *NASA SSP 30425 Rev. B*, 1993.

[2.4.2] Cour-Palais, B.G. Meteoroid Environment Model – 1969. *NASA SP-8013*, NASA Space Vehicle Design Criteria, 1969.

[2.4.3] Grün, E., Zook, H.A., Fechtig, H. *et al.* Collisional Balance of the Meteoritic Complex. *Icarus*, **62**, 244–272, 1985.

[2.4.4] Oswald, M. *et al.* The MASTER 2005 Model. Proceedings of the 4th European Conference on Space Debris, *ESA SP-587*, pp. 235–242, 2005.

[2.4.5] Devine, E., Grün, E., Staubach, P. Modeling the Meteoroid Distributions in Interplanetary Space and Near Earth. Proceedings of the 1st European Conference on Space Debris, *ESA SD-01*, pp. 245–250, 1993.

[2.4.6] Staubach, P. Upgrade of the DISCOS Meteoroid Model. *Final Report, ESA Contract no. 10463/93/D/CS*, Darmstadt, Germany, 1996.

[2.4.7] Rex, D. Die mögliche Überfüllung erdnaher Umlaufbahnen durch die Raumfahrt. In *Carolo-Wilhemina Mitteilungen II/1996 der Technischen Universität Braunschweig*, 1996.

[2.4.8] Chamberlain, S.A., Slauenwhite, T.A. United States Space Command Surveillance Network Overview. Proceedings of the 1st European Conference on Space Debris, *ESA SD-01*, pp. 37–42, 1993.

[2.4.9] Anonymous. Orbital Debris Mitigation Re-emphasized in the New US National Space Policy. *NASA Orbital Debris Quarterly News*, **11** (1), January 2007.

[2.4.10] Klinkrad, H. (ed.) *ESA Space Debris Mitigation Handbook*, Second Edition, Issue 1.0, 2003.

[2.4.11] Liou, J.-Ch. *et al.* The New NASA Orbital Debris Engineering Model ORDEM2000. *NASA/TP-2002-210780*, 2002.

[2.4.12] Eichler, P., Rex, D. Debris Chain Reactions AIAA-90-1365. AIAA/NASA/DOD Orbital Debris Conference: Technical Issues & Future Directions, Baltimore, MD, April 1990.

[2.4.13] Rex, D. The Role of the Scientific and Technical Subcommittee of the UN-COPUOS for the Space Debris Work of the United Nations. Proceedings of the 2nd European Conference on Space Debris, *ESA SP-393*, pp. 759–762, 1997.

[2.4.14] Kessler, D.J. *et al.* Orbital Debris Environment for Spacecraft Designed to Operate in Low Earth Orbit. *NASA TM-100471*, 1989.

[2.4.15] Alby, F., Lansard, E., Michal, T. Collision of CERISE with Space Debris. Proceedings of the 2nd European Conference on Space Debris, *ESA SP-393*, pp. 589–596, 1997.

[2.4.16] Kinslow, R. (ed.) *High-Velocity Impact Phenomena*. New York: Academic Press, 1970.

[2.4.17] Wipple, F.L. Meteorites and Space Travel. *Astron. J.*, **52**, 131, 1947.

[2.4.18] Cour-Palais, B.G. Space Vehicle Meteoroid Shielding Design. Comet Halley Micro Meteroid Hazard Workshop, *ESA SP-153*, pp. 85–92, 1979.

[2.4.19] Gehring, J.W. *et al.* Experimental Studies Concerning the Meteoroid Hazard to Aerospace Material and Structures. *J. Spacecr.*, **2**, 731–737, 1965.

[2.4.20] Christiansen, E.L. Design and Performance Equations for Advanced Meteoroid and Debris Shields. *Int. J. Impact Eng.*, **14**, 145–156, 1993.

[2.4.21] Reimerdes, H.-G., Stecher, K.-H., Lambert, M. Ballistic Limit Equations for the Columbus Double-bumper Shield Concept. Proceedings of the 1st European Conference on Space Debris, *ESA SD-01*, pp. 433–439, 1993.

[2.4.22] Christiansen, E.L., Kerr, J.H. Mesh Double Bumper Shield: A Low-weight Alternative for Spacecraft Meteoroid and Orbital Debris Protection. *Int. J. Impact Eng.*, **17**, 477–486, 1995.

[2.4.23] Destefanis, R., Faraud, M. Testing of Advanced Materials for High Resistance Debris Shielding. *Int. J. Impact Eng.*, **20**, 209–222, 1997.

[2.4.24] Turner, R.J. *et al.* Cost-effective Debris Shields for Unmanned Spacecraft. *Int. J. Impact Eng.*, **26**, 785–796, 2002.

[2.4.25] Schäfer, F. *et al.* The Inter-Agency Space Debris Coordination Committee (IADC) Protection Manual. Proceedings of the 4th European Conference on Space Debris, *ESA SP-587*, pp. 39–46, 2005.

[2.4.26] Cour-Palais, B.G. Hypervelocity Impact Investigations and Meteoroid Shielding Experience Related to Apollo and Skylab. *NASA Conference Publication 2360, Orbital Debris*, pp. 247–275, 1982.

[2.4.27] Hyde, J.L. *et al.* A History of Meteoroid and Orbital Debris Impacts on the Space Shuttle. Proceedings of the 3rd European Conference on Space Debris, *ESA SP-473*, pp. 191–196, 2001.

[2.4.28] Nagao, Y. *et al.* Hypervelocity Impact Studies Simulating Debris Collisions on Composite Materials. Proceedings of the 4th European Conference on Space Debris, *ESA SP-587*, pp. 413–418, 2005.

[2.4.29] Reimerdes, H.-G., Nölke, D., Schäfer, F. Modified Cour-Palais/Christiansen Damage Equations for Double-wall Structures. *Int. J. Impact Eng.*, **33**, 645–654, 2006.

[2.4.30] Wohlers, W., Reimerdes, H.-G. Analytical Optimisation of Protection Systems. *Int. J. Impact Eng.*, **29**, 803–819, 2003.

[2.4.31] Taylor, E.A. *et al.* Hypervelocity Impact on Spacecraft Honeycomb: Hydrocode Simulation and Damage Law. *Int. J. Impact Eng.*, **29**, 691–702, 2003.

[2.4.32] Hiermaier, S.J., Schäfer, F. Simulation of Ellipsoidal Projectile Impact on Whipple Shields. *Int. J. Impact Eng.*, **29**, 333–343, 2003.

[2.4.33] Fahrenthold, E.P., Park, Y.-K. Extension and Validation of a Hybrid Particle–Finite Element Method for Hypervelocity Impact Simulation. *Int. J. Impact Eng.*, **29**, 237–246, 2003.



3 Space Transportation Systems

Horst Holsten

Scientific research and commercial space utilization require transportation systems which carry spacecraft and their payloads into Earth orbit and beyond. The physical basis of all launch vehicles used today is the **repulsion principle**, according to which any change of motion is produced by corresponding opposed forces. This forward impulse, also called thrust, is developed by rocket engines.

By following this principle, the construction of a launch vehicle can be regarded as simple, but seen from the standpoint of technical implementation it is quite complex. Its essential components are propellant tanks, engines, structures, navigation and guidance equipment, separation systems, attitude control systems and flight data monitoring systems. The two propellants, the fuel and the oxidizer, are ignited in the thrust chamber where they produce very high-temperature gases by chemical reactions and then expand through a nozzle, thereby generating thrust.

For **propulsion**, solid or liquid propellants are used, in special cases also in combination. In the case of solid propellants, the combustion components are mixed and cast in a cylindrical mold. Once ignited, the whole block burns continuously; interruption of the combustion process is not possible. In the case of liquid propellants, the fuel and oxidizer are fed separately into the combustion chamber, where they undergo chemical reaction. The flow can be controlled or even interrupted with the possibility of reignition.

Space transportation systems were developed in the second half of the twentieth century to explore extraterrestrial space, the Moon and the planets of the Solar System. Soon Earth-related applications for satellites also started, including weather forecasting, communications, Earth observation and human space flight, culminating in 1969 with the first landing on the Moon.

Launch vehicles have been developed from a number of different motivations. First, there are the national strategic and sovereign aspects, such as achieving autonomous access to space. But market forecasts as well as the technological standards of industry and research projects are also decisive for the conception of space transportation systems. The launch vehicles currently used meet the demand for worldwide **spacecraft transport**. They have different payload capacities regarding mass and volume and are offered at different launch costs. On average, in the present decade 20 satellites in the class of 3–5 tonnes are launched annually. To this sum approximately the same number of small satellites in the class up to 1 tonne can be added. The **transport costs** are in the order of 15 000 euros (20000 US dollars) per kilogram of payload; the reliability of reaching the predicted orbit is higher than 95%.

For future space programs such as **lunar and Mars exploration**, the transportation system is no longer regarded as the critical technological element, presuming that these programs can be executed with well-tested and qualified launch vehicles. Except for

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9

the Space Shuttle, which will be discontinued after 2010, today's vehicles are not designed as reusable vehicles qualified for human space flight, but are one-way or expendable vehicles. For each launch, a new vehicle is necessary. This results in two challenges for future space transportation systems: first, reusability and, second, human space transport qualification, with possibly a combination of both requirements. Basic research in transportation systems has been going on for many years, particularly in the fields of take-off and landing capability, reentry, mass reduction and reusability. Technological breakthroughs can, however, only be expected in the case of drastically increased demand for space applications.

The sections of this chapter will give a detailed introduction to the theory, technology and design of present and future space transportation systems, launch vehicle stages and subsystems. The propulsion requirements for various space missions, different propulsion systems, as well as the development and qualification philosophies for components and complete transportation systems, will be described. Finally, requirements and implementation of launch sites and ground facilities will also be treated.

3.1 Systems

Jens Lassmann and Michael H. Obersteiner

The requirements for **launch systems** (rockets) deviate noticeably from those for other land, water or air vehicles. A substantial difference is that today's systems are not reusable, or reusable only to a very low extent. Even for the partially reusable Space Shuttle system, only the orbiter is refurbished for reuse, and it amounts to only 3% of the take-off mass. The reason is that, compared to other transport systems, the performance requirements for the chosen launch system are extremely high and the partially expendable solution is the only feasible one based on today's technology and experience.

These **extreme requirements** will become clear after the basic principles are presented in Section 3.1.2, following the introduction. The "building blocks" of a rocket are presented thereafter to give an impression of the complexity of the overall system and details of the resulting requirements. Because of the extreme requirements and the complexity of the system, the development of launch systems is an extensive process, in terms of both time and expense. A description of the logistics involved will complete the description, supported by an overview of today's launch systems. At the end future prospects for reusable systems will be given.

3.1.1 Introduction

Launch systems transport all types of spacecraft, which are simply named payload among the rocket community, to destinations in space targets. These destinations can be trajectories or selected places on the surface of planets or other cosmic bodies. In accord with cosmic velocities, launch systems must obtain high speeds and the accelerated payloads are therefore comparatively small. For launch into a low Earth orbit the payload must be accelerated to almost 8 km/s and in this case it amounts to typically 1–5% of the take-off mass of the launch system.

Payloads of a rocket can be of very different types, depending on the particular mission, which may involve:

- Scientific probes and satellites
- Application satellites for
  - Telecommunications
  - Earth observation and meteorology
  - Navigation
- Systems for human space missions.

Since the first satellite launch – Sputnik 1 launched by the USSR on October 4, 1957 – the number of missions per year accomplished worldwide by space launchers has changed noticeably. In the beginning, military Earth observation as well as national prestige projects were of great importance and within 10 years the **number of launches** steadily increased to about 140 per year (including launch failures). Afterwards this trend reversed, mainly because of the introduction of digital image recording and transmission. This continuous reduction led to today's number of

about 60 launches per year. If one examines the different space mission activities and current launcher developments, the trend is predicted to change again to an increasing number of launches.

Today the launch of **telecommunication satellites** into geostationary orbit and/or into transfer orbits plays the most important role, using 30–40% of the available annual launch capacity. Most of these satellites are commercially operated and the costs connected with the launch and the associated risk are an important element in the economic business plan.

Human missions are conducted by the USA (Space Shuttle), Russia and China. With under 10 missions per year they are less significant regarding sheer number, but are of great importance for national budgets and industrial development because of the high expenditures for preparation and execution.

It is becoming increasingly difficult to distinguish between **military and civilian missions**. On one side the military exploit commercial satellites, at least for telecommunications, while on the other side purely military tasks are combined with tasks such as civil security and disaster management. So far active weapon systems have not been placed in Earth orbits, or at least only to a very low extent. Because it is comparatively simple to detect and destroy satellites, it is expected that this will not fundamentally change.

Determined by the payloads, launch systems accomplish different missions. Launch into a low Earth orbit (LEO) is the first step. With a combination of propelled and ballistic (coasting) flight phases the higher orbits necessary for reaching the Moon or interplanetary targets can subsequently be achieved. Interstellar missions (with reasonable mission durations) are beyond the capabilities of today's launch systems. For this, new propulsion technologies would be necessary. Space probes launched in the 1970s (Pioneer and Voyager) – using rocket-powered launch systems and gravitational maneuvers at planets – progressed furthest into outer space. Voyager 1 has reached a distance of 100 astronomical units (i.e., 100 times the distance between the Sun and the Earth) or about 14 light-hours after approximately 30 years of flight time. Proxima Centauri, the closest neighbor of our Sun at about 4 light-years, is thus still very far away.

Closer to home, flight toward our own Sun has its limits too. Realistically, even with a combination of a manageable number of stages and a reasonable payload it is not possible to reach the Sun directly. The Earth's speed on its track around the Sun of 29.8 km/s is extremely high and at present an insurmountable obstacle.

Launch systems are complex machines, consisting typically of several stages which often use varying propulsion systems. The development and required tests are accordingly complex. Extensive development budgets and long development times are necessary. In Europe a multitude of countries and companies work on the development and production of the Ariane and Vega rockets. Political will is almost always the starting point for the decision to develop a launch system. Development and testing of launch systems are accomplished worldwide with government funding. The US company SpaceX is one of the few exceptions. SpaceX is about to make the Falcon booster rocket operational with essentially only private investment. However, support in the form of government-funded technologies and infrastructure was also necessary.

The political will for government funding is also motivated by the policy to support the national space industry with technologically interesting and challenging tasks. Therefore, as part of European cooperation the so-called **georeturn rule** applies, ensuring the return of an ESA member state's financial contributions back to that member state's national industry.

The industrial constellations have likewise changed since the first space launches and the trend to introduce more commercialization and rationalization is unmistakable. A limited number of system companies cooperate with multiple suppliers. The most important component or the most important subsystem of booster rockets and/or their stages is the propulsion system. The manufacturers of civil solid propellant engines are typically also manufacturers of military rocket propulsion systems. The manufacturers of liquid propellant rocket systems are mostly united with aviation engine manufacturers. About 7000 people altogether are employed in developing and operating launch systems in Europe. A small number of them, about 200 engineers, work at **Arianespace**, the European commercial operator

company of the Ariane rocket. In the future this company will also operate the new smaller launch vehicle, Vega. In addition, Arianespace, in cooperation with a Russian manufacturing firm, will launch Soyuz rockets from the European launch site of Kourou in South America, presumably starting from 2009.

Arianespace was the first **launch operator** (also called launch service provider) to concentrate on the launch of commercial satellites. Success helped Europe to keep the institutional spending necessary for the infrastructure – often also called fixed costs – attractively low. The infrastructure for launcher production and operation comprises the launch pad, trajectory tracking, weather forecasting and data communication. In addition, special industrial infrastructure for manufacturing, integration and testing must be included.

As a result of the historical evolutionary process, fixed and development costs are financed by government budgets. Driven by competing launcher systems and low prices (in particular Russian manufacturers offer extremely low prices) the world market prices for commercial space launcher systems dropped so far that Arianespace could no longer recover a substantial portion of the fixed costs. But after US manufacturers withdrew from the commercial market and the Russian launchers increasingly faced problems of quality and costing, price development has tended again toward cost coverage, including the fixed costs.

3.1.2 Fundamentals

To understand launch systems, as a first step one has to understand the basic aspects of the task and design of a launcher. In principle the task is to launch a payload into its designated orbit, and computation of the performance of the launcher is the basis of the design. Ascent against Earth's gravitational force to reach an Earth orbit is the primary task of a space launch system. For today's usually rocket-powered launch vehicles, the fundamental rocket equation in combination with the principle of staging provides a necessary, but not sufficient, basis for assessing performance.

3.1.2.1 Payloads and Missions

Knowing the payload mass and the target orbit a simple **mission analysis** can be accomplished. The core of this simple mission analysis is to determine the launch system's capability of producing the required velocity change Δv necessary to execute the mission.

This velocity increment is determined to a large extent by the velocity of the target orbit. Other considerations are:

- The initial speed due to the Earth's rotation
- The gravitational losses
- The losses from air drag
- Steering losses.

Additional details will be given in the following sections.

The **velocity gain** provided by the launch system must at least correspond to the velocity requirement of the mission. Computation of the ideal velocity gain of the launch system using the fundamental rocket equation needs as input:

- Payload mass
- Number of stages
- Propulsion performance (for each stage)
- Structural mass and fuel residuals (for each stage).

The derivation of the theoretical basis and the associated formulas are likewise given in the following sections.

A further important component of the mission analysis is the introduction of **margins**. A simple, but success-promising logic is: the simpler the estimation, the more margins must be introduced. Justified by the uncertainties of the estimations, margins can be defined for:

- Payload mass
- Velocity requirement of the mission
- Propulsion performance (for each stage)
- Structural mass (for each stage).

The margins can be reduced by improving the calculation of performance with the help of a detailed simulation of the launch system. Typically these numeric simulations are mixtures of theoretically derived equations (from mechanics, thermodynamics, aerodynamics, reaction chemistry, etc.) and empirical

relations combined with actual experience. Examples of such experience are included in the following sections.

3.1.2.2 Momentum and the Fundamental Rocket Equation

Rockets use propellant, which they carry along, to produce thrust and velocity. With this principle they are to a large extent independent of their environment and therefore suitable for traveling into outer space. On the basis of the principle of the conservation of linear momentum, the discharging of an incremental particle with mass dm and speed c causes an opposite momentum on the rocket, that is a speed change dv of the rocket mass m:

$$m \cdot \mathrm{d}v = \mathrm{c} \cdot \mathrm{d}m \tag{3.1.1}$$

If the exhaust velocity c is constant and if no further forces affect the rocket mass, then the variables can be separated and integrated for the fundamental rocket equation, which was conceived in 1903 by the Russian scientist K. Tsiolkovsky (or Ziolkowski):

$$v = c \cdot \ln \frac{m_0}{m_e} \quad [\mathrm{m/s}] \tag{3.1.2}$$

where:

v = ideal velocity gain,
m_0 = initial mass of the rocket,
m_e = mass at burnout.

The demand for a constant exhaust velocity is fulfilled to a good approximation by chemical rocket propulsion, which is used in today's launch systems.

From the principle of the conservation of linear momentum the **thrust** F of the rocket also can be derived as the product of the mass flow $\dot{m}$ and the exhaust velocity c:

$$F = \dot{m} \cdot \mathrm{c} \ [\mathrm{N}] \tag{3.1.3}$$

From the thrust and the rocket mass the acceleration a can be computed:

$$a = F / m \ [\mathrm{m/s^2}] \tag{3.1.4}$$

All considerations of forces and acceleration must take into account that thrust, air drag, lift and similar kinds of forces are contact forces. The force of gravity is by contrast an interior force, and its effect on the reaction forces within a rocket and on its movement must therefore be regarded differently.

3.1.2.3 Staging

A single rocket stage can hardly reach orbital speed unless the payload is very small. This limitation can be overcome by staging. The ideal velocity gain of the individual stages then adds up to the **total velocity gain** v_G:

$$v_G = v_1 + v_2 + v_3 \ldots \tag{3.1.5}$$

Computation of the total velocity gain based on the fundamental rocket equation requires consideration of the masses of the upper stages in the initial and burnout masses of the lower stages. The velocity gain of the individual stages is thereby reduced. However, in total a larger velocity gain and/or larger payload performance is achieved. But this increase is limited. The maximum performance of a staged rocket launcher results to a first approximation in the following **number of stages**:

$$\boldsymbol{n}_{\mathrm{opt}} = 1.12 \cdot v/\bar{c} \tag{3.1.6}$$

where $\bar{c}$ is the average value of the exhaust velocities of all stages. As the equation delivers typically a decimal number it must be rounded to the next integer.

With the above equations the efficiency of launch vehicles can be computed as a first approximation. However, reference values for the employed parameters are necessary.

To estimate the **exhaust velocity** three characteristic propellant groups can be distinguished:

- Solid propellant engines, $c \sim 2500$ m/s
- Medium-energy liquid propulsion, $c \sim 3000$ m/s
- High-energy liquid propulsion, $c \sim 4000$ m/s.

From this the **specific impulse** results from the equation

$$I_{sp} = c/g_0 \ [\mathrm{s}] \tag{3.1.7}$$

as a common performance figure for propulsion.

For the computation of the mass ratios it is necessary to estimate the **structural mass**. It varies typically between 5 and 15% of the total mass depending on the size of the rocket stage and the propulsion and/or propellants used.

It is meaningful to use solid propellant engines for the first flight phase and high-energy liquid propulsion for the upper stages. The solid propellant stages are well suited to produce the take-off thrust. The high-energy and accordingly more efficient upper stages are less heavy and thus reduce the propulsion task of the lower stages. Therefore the investment in high-energy propellant upper stages also results in a reduction of the mass of the lower stages.

3.1.2.4 Ascent Trajectory and Performance Requirements

The ascent of a launch vehicle begins **vertically**. This is due to the technical feasibility and the dynamic behavior during the launch phase. For a successful take-off it is necessary for the thrust to be at least equal to the launch vehicle mass. A 10–20% higher take-off thrust helps the launcher, which is flight dynamically unstable, to maintain both the correct flight attitude by swiveling the rocket engines and the acceleration to build up the required speed. However, further increasing the take-off thrust is of no advantage. On the one hand additional investments in propulsion would become necessary, and on the other hand the acceleration at burnout would become even higher, putting additional strain on the system, because when the major part of the propellant is used up, the lighter weight of the rocket experiences high acceleration.

To reach a horizontal orbit it is necessary to **tilt** the rocket shortly after take-off by roughly 10°. After the tilt maneuver the thrust direction can be optimally oriented so that it is tangential to the flight direction. Gravitational force then turns the trajectory toward a horizontal orientation. The initial tilt must be matched with the acceleration level and the target orbit in such a way that at burnout the desired orbital speed and flight attitude are reached.

The **velocity requirement** for space launch systems results mainly from the difference between the initial and terminal speeds. For the initial speed the Earth's rotation must be taken into account; it of course depends on the geographical latitude of the launch site and the direction of the trajectory. The terminal velocity is the target orbital speed.

The ascent from the surface to an orbit is an ascent inside Earth's gravitational field. Thus so-called **gravitational losses** will occur. This corresponds not only to the gravitational energy invested in the payload, but also to the gravitational energy invested in the sections of the launch vehicle to be transported upward, and in particular in the build-up of the "fuel and/or exhaust tower." Coasting flight phases in the out-of-atmosphere part of the ascent of launch vehicles can help to reduce gravitational losses. **Air drag and steering losses** lead to additional velocity requirements. At the end, margins for unexpected orbit/trajectory disturbances and inaccuracies must be included.

The velocity requirement for reaching a low Earth orbit (~200 km altitude) adds up as follows:

- Orbital velocity: 7.9 km/s
- Earth rotation speed gains at launch: 0–465 m/s
- Gravitational losses: 1–1.5 km/s
- Air drag losses: 100–150 m/s
- Steering losses: 20–50 m/s
- Performance margin: 1–2%
- **Total:** 8.6–9.8 km/s.

Missions to a Sun-synchronous orbit (SSO), geostationary orbit (GEO) and/or its transfer orbit (GTO), lunar missions or escape missions from Earth's gravitational field result in accordingly higher velocity requirements. Values for the launch from the European launch site at Kourou in South America are given in Table 3.1.1.

Table 3.1.1: *Velocity requirements for various target orbits with launch from Kourou.*

Target orbit	Inclination	Velocity requirement [km/s]
LEO 200 km	5°	9
SSO 700 km	98°	9.7
GTO 200/36 000 km	0° (equatorial)	11.6
Moon (impact)		12.0
Escape		12.5
Lunar orbit		≈13
Mars orbit		≈15

3.1.3 Building Blocks

A launch system consists of a launch vehicle and a ground segment and is composed of a set of building blocks whose main task is to transfer payloads into space.

The launch vehicle itself consists of the following building blocks:

- Boost stages
- Propulsive stages including engines
- Equipment bay
- Payload adaptors
- Multiple payload launch system
- Payload fairings.

Boost stages with their high-thrust engines provide additional accelerating power during the first flight phase for reducing large gravitational losses. Mainly solid propulsion and medium-energy liquid propellant rocket engines are used.

In general two, three or four stages form the propulsive system of a launch vehicle. Each of these propulsive stages consists of propellant tanks, propellant pumps, rocket engines, attitude control system and an interstage including the stage separation system. Propellants and propulsive system concepts may differ for each stage. A performance-optimized launch vehicle is made of high-thrust lower stages and an upper stage with a high specific impulse engine.

Rocket stages can be bundled in parallel, as is usually the case for boost stages and the first stage, or they are operated in series, that is they are ignited sequentially.

Guidance, navigation and control of a launch vehicle are fully autonomous during the flight. This is achieved via individual avionics systems within each stage (e.g., Soyuz) or centrally by the upper stage for the whole launcher (e.g., Ariane 5).

The **payload** is housed in the so-called payload compartment, consisting of the payload adaptors, the multiple payload launch system and the payload fairings. Payload adaptors are the structural interface between the launch vehicle and the payload, which usually has a smaller diameter than the rocket itself. Large launch vehicles like Ariane 5 are capable of launching multiple payloads independent of each other during the same flight. To do this they use specific multiple payload launch systems or payload dispensers. This technique leads to attractive launch prices per payload.

The launch vehicle alone is not capable of launching a payload. Without the **ground segment** and its specific infrastructure a launch vehicle cannot be operated. This ground infrastructure consists mainly of:

- Launch vehicle integration buildings
- Payload preparation buildings
- Launch platform
- Launch zone including launch tower, lightning masts, exhaust ducts, propellant, gases and water provision
- Control facilities for launcher and payloads
- Telemetry and tracking stations along the flight path of the launcher
- Radar systems
- Facilities for weather forecasting
- Production facilities for cryogenic propellants.

In addition, the launch site must provide adequate infrastructure for delivering and storing the launcher stages, payloads, propellants and gases. Operational personnel and customer representatives must be accommodated nearby. This means that not only hotels, restaurants and grocery stores have to be available, but also schools and other social facilities.

3.1.4 Project Phases

A typical launch system project can be split into several phases of its life cycle. In order to start a launcher development project all the enabling technologies have to be mature enough. The maturity of analytical methods, materials and manufacturing processes can be described by the **technology readiness level** (TRL) and should be at least at level 4 to 6 when starting a development. The development of technologies can be a lengthy process before they are ready to be applied. As an example, it took about 10 years to develop an aluminum alloy that can be welded for manufacturing the cryogenic oxygen tank for Ariane.

The **concept phase** as the first step of launch vehicle development starts independently of the maturity of the technologies and is mainly for finding ideas and preparing the development program. Based on

various concepts a program proposal is established which leads to an invitation to tender issued by space agencies in the case of public funding. The duration of the concept phase is linked to the decision cycles of the public customers and typically lasts for four to ten years.

Work on concrete launcher configurations starts during the **preliminary design phase**. Starting with high-level functional requirements, the launch system is defined and different technical solutions are subjected to trade-off investigations. This phase takes typically 20% of the development time and is concluded by the **preliminary design review** (PDR). The definition of the launcher and its subsystems and the analytical justification for their functioning are part of the design phase. Development tests for calibrating mathematical models may accompany these activities. This second concrete **development** step takes about 30% of the development time and is concluded by the **critical design review** (CDR).

The following step qualifies the launch system on the ground. Tests are performed to verify the mechanical integrity as well as to demonstrate the functioning of the engines and the hydraulic and electrical systems. These tests are carried out on component, subsystem and system levels. Based on these tests all mathematical models are validated in order to allow precise prediction of the launcher's behavior on the ground and in flight. This phase takes about 40% of the development time and is completed by the **qualification commission**. As evidence of their agreement on the flightworthiness of the launcher, all involved parties (customer, certifying agency and development authority/contractor) sign the qualification certificate.

In parallel to the qualification, the **manufacturing** process for the first flight system hardware begins. Following the first launch an intensive evaluation of the data measured in flight is made. The comparison of predicted values to the measured data leads to a final correlation with the mathematical models. This enables an analysis of potential deviations during the production phase. Thus it is possible to fully justify decisions to use or to delete hardware items. Typically small changes to the launcher are introduced after this phase in order to correct behavior deviations and to achieve a sufficient robustness of the design for the series production. Including the first flight, this last phase takes about 10% of the development time. In total, a laŭnch system development requires about 10 years, starting with the preliminary design and ending with series production.

During the **operations phase** the configuration of the launcher should be kept frozen. Only a minimum number of changes should be made in order not to deviate from the qualified status. Besides this consideration, the stability of the manufacturing processes contributes to a large extent to the reliability of a launch system. An accompanying test program is often performed in parallel to the series production in order to verify that the qualification of critical components like engines is being maintained.

At the **end of the operations phase** the production of the launcher and its components is terminated. Consequently a sufficient number of spare parts have to be available also for the last launcher in order not to endanger the mission in case of integration problems. This run-down phase is further penalized by a decrease in awareness on the part of the manufacturing personnel since those who are well experienced will already have moved on to new projects. In order to mitigate the risk of mission loss, the quality assurance effort should be significantly increased during this last phase of the launch system life cycle (Figure 3.1.1).

3.1.5 Overview of Launch Systems

Launch systems are the key to space and therefore of high strategic interest. Today the following nations operate launch systems for civil use: Russia and Ukraine (since 1957), USA (since 1958), Europe (since 1965), Japan (since 1970), China (since 1970), India (since 1980) and Israel (since 1990). Other launch system development activities are ongoing in Brazil, Korea and Iran.

There are various ways to categorize launch systems. According to the number of stages, one distinguishes **single stage to orbit** (SSTO) from **multiple stage** launchers (mostly two stages to LEO or three stages to GTO). Launchers can be categorized as **expendable launch vehicles** (ELVs) or as fully or partly **reusable launch vehicles** (RLVs). All of these launcher types can be **launched vertically**

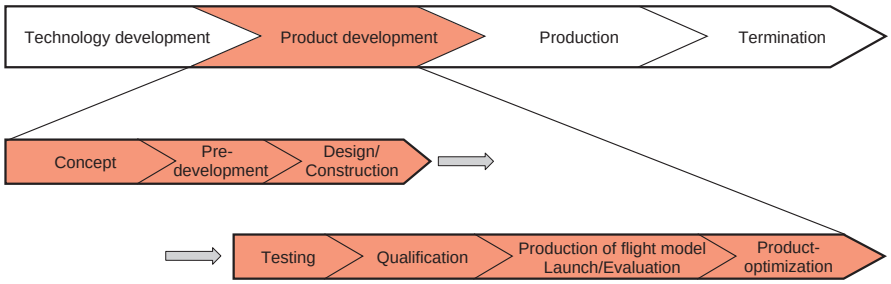


Figure 3.1.1: *Life cycle of a launch system project.*

(VTO: Vertical Take-Off); winged launchers can be **launched horizontally** (HTO: Horizontal Take-Off). Further, all types of launchers can be qualified for human space flight. Those then need a high reliability which must be implemented in the vehicle design and a return module. Another often-used categorization is to distinguish launchers for commercial use from "institutional" launchers.

3.1.5.1 Present Launch Systems

The majority of spacecraft launched today are satellites and automated scientific space probes. In the following, launch systems are therefore categorized according to how much payload mass is placed into the GEO transfer (GTO) orbit: small launchers up to 3000 kg, medium-sized launchers between 3000 and 6000 kg, and heavy-lift launchers with more than 6000 kg.

Small launchers are mainly used for transferring payloads into LEO. Small launchers are usually the first step to gain access to space due to their lower technical complexity. Nations that are not one of the pioneers in space launches but declare an independent access to space as a national strategic priority are first successful with small launchers: Israel with Shavit, India with PSLV and GSLV, and Italy with Vega are examples. Commercial launch system initiatives follow the same principle, as demonstrated by Orbital Sciences with Pegasus XL and Minotaur, and SpaceX with the Falcon launcher. In addition, a number of small launch systems are operated on the basis of former intercontinental ballistic missiles. Utilizing former weapon systems for space exploration became possible though international treaties for reducing strategic arsenals. It concerns currently only the former Soviet systems like Rockot, Cosmos-3M, Dnepr, Shtil', Start-1 and Tsyklon-1. The Chinese small-launcher version of "Long March" (CZ-4) is also based on the building blocks of ballistic missiles.

Medium launchers serve to place satellites into all Earth orbits: LEO, including polar orbits, medium Earth orbits (MEOs), GTOs, GEO and Earth escape missions. Even though they have been operated for decades, medium-sized launchers are still kept operational in parallel to their successors with larger payload capacity mainly because they are optimized for a specific type of mission: for example, Soyuz lifts cosmonauts into LEO, aside from the cargo version; Delta II, the new Delta IV-M and the Chinese "Long March" versions CZ-2 and CZ-3 serve only nationally protected institutional markets. In addition, the Chinese CZ-2F is qualified for crew transport, which allows China as the third nation after Russia and the USA to launch people into space and safely return them to Earth. Sea Launch using Zenit-SL and Arianespace using Soyuz launch vehicles are examples of the commercial marketing of medium-sized launcher systems. The Japanese H-IIA is one of the few examples of a medium-sized launcher for a newly developed second-generation system following a small-launcher development.

Heavy-lift launch vehicles (HLLVs) mainly launch communications satellites into GTOs and are used specifically for launching very heavy payloads. HLLVs dominate the commercial market for satellite launches. Since the founding of Arianespace in 1980, space transportation has become commercial. Initially, all of today's existing launchers were used for commercial launches. Competition, especially with nations of the former Soviet Union, limits commercial space transportation today to the Ariane 5, Proton, Atlas V and the Japanese H-IIA-204. Delta IV-H and STS (the Space

Shuttle) are used solely for institutional launches. The transport of commercial payloads using the Space Shuttle was discontinued in 1986 after the loss of the *Challenger* orbiter. As the most important strategic launch system of the USA and the only one capable of launching crews, the STS is thereby protected from any risk of loss of mission when launching commercial satellites.

Some representative current launcher systems are presented next.

Rockot (Figure 3.1.2, Table 3.1.2) is a launch system based on the former SS-19 intercontinental ballistic missile. The initially two-stage missile with its storable propulsion systems was upgraded with an autonomous storable liquid-propelled upper stage Breeze-KM and a new telemetry system. Rockot is launched from the Russian space port of Plesetsk mainly into polar LEOs. The upper stage Breeze-KM is reignitable and thus able to lift multiple payloads into different orbits with one launch. Rockot is commercially available and is being marketed by the German–Russian joint venture EUROCKOT.



Figure 3.1.2: *Rockot (Source: EUROCKOT).*

Pegasus-XL (Figure 3.1.3, Table 3.1.2) is a three-stage launch system using solid propulsion and is air launched from a modified civil aircraft. This makes it possible to reduce the required thrust level at liftoff, and because of the higher elevation of the "launch pad", a longer nozzle extension can be used to improve the specific impulse of the first stage. In addition an aircraft as "launch pad" is geographically highly flexible and enables optimized injection into the desired inclination of the target orbit. A strategic role is attributed to Pegasus for quickly launching small surveillance satellites when immediately needed by the USA. Pegasus is only used for launching institutional payloads since its comparably high launch cost is not competitive.

Falcon 1 (Figure 3.1.4, Table 3.1.2) is the first fully commercial development of a new launch system. Financed to a large extent by the founder of SpaceX, it is subsidized by the first launch contracts placed by the US government. Falcon 1 was launched successfully for the first time in 2008; it is a two-stage launcher using semicryogenic propellants (LOX/kerosene). The launch site is at Kwajalein, a US military missile test site in the South Pacific.

Vega (Figure 3.1.5, Table 3.1.2) is another example of a newly developed small launcher. Initiated by the Italian Space Agency (ASI), Vega became an ESA development program. Italy makes use of the launch system experience gained in the 1980s when Scout rockets manufactured under license from the USA



Figure 3.1.3: *Pegasus XL mounted under the launching aircraft (Source: OSC).*



Figure 3.1.4: *Falcon 1 (Source: SpaceX).*

were launched from San Marco off the coast of Kenya. Vega is a three-stage solid propellant launch system with an orbit injection module using storable liquid propellants. Like Ariane 5, Vega is launched from CSG near Kourou, French Guiana, and is operated commercially by Arianespace.

Soyuz (Figure 3.1.6, Table 3.1.3) is the world's most successful launcher of medium payload capacity. Since 1966 more than 1200 Soyuz rockets have been launched from the Russian launch site at Baikonur. Based on the former Soviet intercontinental ballistic missile R-7, it was developed specifically to launch cosmonauts into LEOs. Its basic version is a 2.5-staged launcher using semicryogenic propellants. Four LOX/kerosene boost stages are attached around a core stage. For launching satellite payloads the Fregat upper stage, which can be reignited up to 20 times, is used. Initially qualified for crew transport, Soyuz is still operational for such missions, utilizing



Figure 3.1.5: *Vega (Source: ESA).*

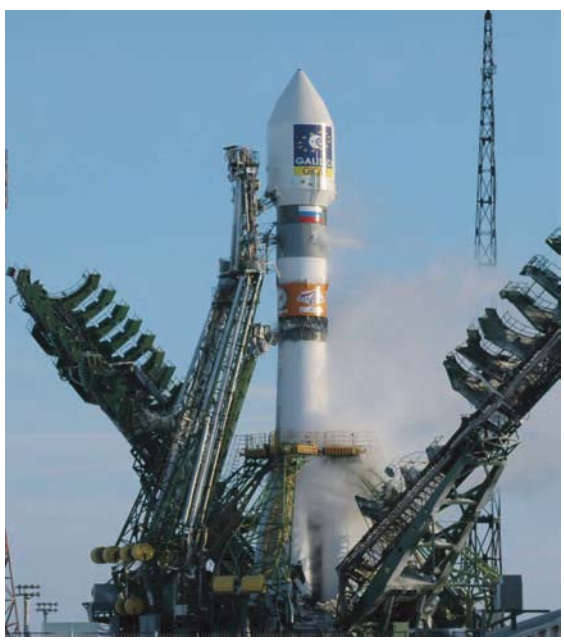


Figure 3.1.6: *Soyuz (Source: ESA).*

the Soyuz crew module as the upper stage. Commercial Soyuz launches are marketed internationally by Arianespace. Starting in 2009, Soyuz will also be launched from the European launch site CSG in French Guiana. A technical specialty is the "hot" stage separation. Differing from most other launchers, the stages are separated while in the thrust decay phase of the first stage, with the second-stage engine already running. Using this principle an additional pre-acceleration system is not needed for orienting the upper stage propellants at the tank outlet prior to engine ignition.

CZ-4B (Chang Zheng, "Long March") (Figure 3.1.7, Table 3.1.3) is the workhorse of Chinese space activities. Primarily used for launching Chinese national satellite payloads, this launcher and the slightly higher performing CZ-3 are available on the commercial market. Export regulations placed on China by several industrial nations heavily restrict the commercial use of "Long March." In addition, commercial availability is limited by the high national launch demand of China. Initially a derivative of the DF-5 intercontinental missile, CZ-4B is a three-stage launcher propelled by storable propulsion. Since 1999 the "Long March" launcher in its CZ-2F version has been qualified for crew transport. For this the upper stage is replaced by a crew orbiting and deorbiting module, and for boost enhancement more stages are added, which also consume storable liquid propellants. Crew missions are launched from the Juiquan launch site, whereas satellite missions are launched from Taiyuan.

Zenit-SL (Figure 3.1.8, Table 3.1.3) is a three-stage launch system using semicryogenic propulsion systems. Initially developed as a booster stage for the Soviet heavy-lift launch system Energia, it was later combined with two upper stages for launching satellites also from the Russian Baikonur launch site. A US/Ukrainian/Norwegian joint venture adapted this launcher so it



Figure 3.1.7: *Long March CZ-4B (Source: CNSA).*



Figure 3.1.8: *Zenit-SL (Source: Boeing).*

could be launched from a modified oil drilling rig. By doing so the energetic disadvantages of US and Russian launch sites for serving equatorial orbits are compensated. Sea Launch operates and offers this launcher, designated Zenit-SL, to commercial customers. Launcher and payload are prepared in Los Angeles (USA) and then transferred via ship to the launch platform near the equator where fueling and launch take place. Zenit-SL, Ariane 5 and Proton share the commercial satellite launch market between them.

H-IIA (Figure 3.1.9, Table 3.1.3) is an upgrade of the Japanese H2 launcher, which was a large technological step for Japan following the H-I. The development of H-2A was driven by the need to drastically reduce launch costs and increase the launcher's availability compared to H-II. H-IIA consists of a cryogenic first stage with two or four solid propellant boost stages and a cryogenic upper stage. The first stage is propelled by a high-pressure staged combustion cycle engine and is therefore one of the technologically most advanced launcher stages that have been realized. The H-IIA is offered globally as a commercial launcher besides serving the Japanese national market.



Figure 3.1.9: *H-IIA (Source: JAXA).*

However, the limited number of launch opportunities and thus the low annual launch rate reduce its commercial attractiveness. Launches from the Tanegashima site in Japan are possible for only two three-week periods each year. In the remaining time the coastal fishing industry must remain undisturbed.

Delta IV-M (Figure 3.1.10, Table 3.1.3) is part of a family of launchers which serves the medium-payload segment as well as heavy payloads in its Delta IV-H configuration. It consists of a newly developed cryogenic first stage that can be combined with two solid boost stages. The cryogenic upper stage is an adaptation of the Centaur rocket which was and still is used in several US launchers. The Delta IV first stage is based on a production cost-optimized cryogenic engine. The Delta IV-H version uses two additional liquid first stages as boosters instead of the two solid boost stages of Delta IV-M. This demonstrates the



Figure 3.1.10: *Delta IV-M (Source: USAF).*

Table 3.1.2: *Small launchers.*

	Rockot	Pegasus XL	Falcon	Vega
First stage	RS-18 ICBM first stage	Orion 50S-XL	Liquid propellant stage	P80 FW
Length/diameter	17.2 m/2.5 m	10.27 m/1.28 m	n.a./1.678 m	7.7 m/3 m
Take-off mass	86 689 kg	16 383 kg	22 388 kg	87 438 kg
Propellant mass	80 994 kg	15 014 kg	21 092 kg	80 000 kg
Propellants	N_2O_4/UDMH	HTPB/AP	LOX/PR-1	HTPB-Al/AP
Burning duration	121 s	68.6 s	196 s	105 s
Main engine	4×RD-0233	Solid motor	1×Merlin 1	Solid motor
Thrust (vac/sl)	580 kN/n.n.	594 kN/486.7 kN	320 kN/318 kN	2092 kN/1440 kN
I_{sp} (vac/sl)	310 s/285 s	295 s/n.n.	304 s/255 s	n.a./280 s
Remarks		Launch from aircraft; winged	Parachute system for recovery	
Second stage	RS-18 ICBM second stage	Orion 50-XL	Liquid propellant stage	Zefiro 23
Length/diameter	3.9 m/2.5 m	3.11 m/1.28 m	n.a./1.678 m	7.5 m/1.897 m
Take-off mass	15 481 kg	4341 kg	3745 kg	25 900 kg
Propellant mass	13 996 kg	3926 kg	3385 kg	24 000 kg
Propellants	N_2O_4/UDMH	HTPB/AP	LOX/PR-1	HTPB/AP
Burning duration	183 s	89.4 s	378 s	71 s
Main engine	1×RD-0235	Solid motor	1×Kestrel	Solid motor
Thrust (vacuum)	240 kN	153 kN	33.6 kN	959 kN
I_{sp} (vacuum)	320 s	289 s	325 s	289 s
Remarks	4×RD-0236 vernier			
Third stage	Breeze-KM	Orion 38		Zefiro P9
Length/diameter	2.9 m/2.5 m	1.34 m/0.97 m		3.6 m/1.9 m
Take-off mass	6475 kg	896 kg		10 500 kg
Propellant mass	5055 kg	77 kg		9500 kg
Propellants	N_2O_4/UDMH	HTPB-Al/AP		HTPB/AP
Burning duration	1000 s	68.5 s		116 s
Main engine	1×S5.98M	Solid motor		Solid motor
Thrust (vacuum)	19.62 kN	34.6 kN		230 kN
I_{sp} (vacuum)	325.5 s	283.7 s		294 s
Remarks	8 ignitions			
Fourth stage				AVUM
Length/diameter				1.8 m/1.9 m
Take-off mass				719 kg
Propellant mass				370 kg
Propellants				Hydrazine/NTO
Burning duration				620 s
Main engine				1×hot gas thruster+ 6×cold gas thrusters
Thrust (vacuum)				2.2 kN
I_{sp} (vacuum)				317 s
Remarks				Attitude control for payload orientation
Payload fairings				
Volume	23.52 m^3	1.95 m^3	3.01 m^3	19.68 m^3
Mass	700 kg	170 kg	n.n.	n.n.
Length/diameter	6.73 m/2.52 m	2.139 m/1.16 m	2.79 m/1.048 m	6.3 m/3.0 m
Payload	1900 kg LEO 90°	345 kg LEO 90°	n.n.	1500 kg LEO 90°

Table 3.1.3: *Medium-capacity launchers.*

	Soyuz FG/Fregat	CZ-4B Long March	Zenit 3-SL	H-IIA	Delta-IVM
Boost stages	4×liquid propellant stage			n.n.	2×GEM-60
Length/diameter	19.6 m/2.68 m				
Take-off mass	43 400 kg				33 789 kg
Propellant mass	39 600 kg				29 920 kg
Propellants	LOX/kerosene RG-1				HTPB/AP
Burning duration	118 s				90 s
Main engine	1 × RD-107A			Solid motor	Solid motor
Thrust (vac/sl)	810 kN				608.1 kN/n.n.
I_{sp} (vac/sl)	319 s/257 s				278 s/273 s
Remarks	Four-chamber engine				
First stage	1×liquid propellant stage	1×liquid propellant stage	1×liquid propellant stage	1×liquid propellant stage	1×CBC
Length/diameter	27.14 m/2.95 m	24.66 m/3.35 m	32.9 m/3.9 m	37.2 m/4.0 m	n.a./5.0 m
Take-off mass	101 900 kg	192 700 kg	353 870 kg	113 600 kg	226 400 kg
Propellant mass	95 400 kg	183 200 kg	325 400 kg	100 000 kg	199 640 kg
Propellants	LOX/kerosene RG-1	UDMH/NTO	LOX/kerosene	LOX/LH_2	LOX/LH_2
Burning duration	286 s	170 s	143 s	397 s	251 s
Main engine	1×RD-108A	1×DaFY6-2	1×RD-171M	1×LE-7A	1×RS-68
Thrust (vac/sl)	780 kN	n.n./2961 kN	7850 kN/n.n.	1080 kN/840 kN	3317 kN/2910 kN
I_{sp} (vac/sl)	319 s/248 s	n.n./260 s	337.2 s/309.5 s	443 s/n.n.	410 s/360 s
Remarks	Four-chamber engine				
Second stage	1×Block L	1×liquid propellant stage	1×liquid propellant stage	1×liquid propellant stage	1×Centaur
Length/diameter	6.7 m/2.66 m	10.41 m/3.35 m	10.4 m/3.9 m	9.2 m/4.0 m	12.2 m/4.0 m
Take-off mass	25 200 kg	39 600 kg	92 473 kg	19 600 kg	24 170 kg
Propellant mass	22 800 kg	35 600 kg	81 850 kg	16 600 kg	21 280 kg
Propellants	LOX/kerosene T-1	UDMH/NTO	LOX/kerosene	LOX/LH_2	LOX/LH_2
Burning duration	240 s	127 s	293 s	534 s	850 s
Main engine	1×RD-0110	1×YF-22	1×RD-120	1×LE-5B	1×RL-10B-2
Thrust (vacuum)	297.9 kN	742 kN	833 kN	141 kN	110 kN
I_{sp} (vacuum)	325 s	296.7 s	350 s	447 s	466 s
Vernier engine		4×YF23	4×RD-8		
Thrust (vacuum)			78 kN		
I_{sp} (vacuum)			342 s		
Remarks	Four-chamber engine, five vernier engines		4×RD-8 vernier engines with storable propellants		
Third stage	Fregat	1×liquid propellant stage	Block DM-SL		
Length/diameter	1.5 m/3.35 m	1.92 m/2.9 m	5.6 m/3.7 m		
Take-off mass	6350 kg	15 150 kg	18 350 kg		
Propellant mass	5350 kg	14 152 kg	15 200 kg		
Propellants	UDMH/NTO	UDMH/NTO	LOX/kerosene		
Burning duration	877 s	135 s	660 s		
Main engine	1×S5.92	1×YF-40	1×11D58M		
Thrust (vacuum)	19.6 kN	100 kN	84 kN		
I_{sp} (vacuum)	331 s	306 s	355 s		
Remarks	20 ignitions		7 ignitions		
Payload fairings	Type S				
Mass	1045 kg	n.n.	2000 kg	1716 kg	1677 kg
Length/diameter	5.87 m/3.44 m	6.51 m/3.0 m	8.54 m/4.94 m	9.12 m/4.6 m	10.32 m/4.53 m
Payload	1200 kg GTO	550 kg GTO 4680 kg LEO	6100 kg GTO	4100 kg GTO	5300 kg GTO

high flexibility of this launcher family concept with regard to different payload sizes. Delta IV is operated by Boeing and launched both from the USA's Kennedy Space Center in Florida and from Vandenberg Air Force Base in California.

Ariane 5 (Figure 3.1.11, Table 3.1.4) is a 2.5-stage launch system primarily used for lifting heavy satellites to a GTO. Its capability of launching two satellites totaling a maximum of 9000 kg to GTO is the key to its commercial success. Arianespace as the launch provider has increased Ariane's share of the commercial market to 60% following a drop caused by Russian launch systems entering the market. Two versions are operated today: A5E/CA with a cryogenic upper stage for launching commercial payloads and A5ES-ATV for lifting the ATV supply vehicle to the ISS. The upper stage of this version is reignitable and uses storable liquid propellants. Both versions use a cryogenic first stage and two large solid propellant boost stages. Ariane 5 is launched from the European CSG spaceport near Kourou in French Guiana. Due to the proximity of its launch site to the equator (5°N) Ariane 5 has a clear performance advantage over US and Russian launchers.

Proton (Figure 3.1.12, Table 3.1.4) is, besides Ariane 5, the most important launcher for lifting commercial payloads. Proton is a four-stage launcher operated from the Baikonur launch site in Russia. In contrast to Soyuz-type launchers, Proton is integrated vertically and also transferred vertically to the launch pad. The first three stages are propelled by storable liquids with identical engines that differ only in the length of their nozzle extensions. Alternatively, Block D with a semicryogenic propulsion system or the Breeze upper stage which can be reignited up to 20 times



Figure 3.1.11: *Ariane 5 (Source: ESA).*



Figure 3.1.12: *Proton (Source: NASA).*

can be used as the fourth stage. Proton is commercially offered by the operator ILS.

Atlas V (Figure 3.1.13, Table 3.1.4) is, besides Delta IV-H, the second US heavy-lift launch vehicle that guarantees access to space for US strategic payloads. A Russian multichamber main engine provides the thrust to the semicryogenic first stage. Depending on the needed payload performance, two, four or five solid propellant boosters provide the necessary additional liftoff thrust. As the upper stage, the cryogenic Centaur with one RL-10 engine is used. Even though marketed commercially by Lockheed Martin, Atlas V has a very small market due to its high launch price. Atlas V is operated from the US Kennedy Space Center launch site.

STS (Figure 3.1.14, Table 3.1.4). The Space Transportation System (the Space Shuttle) is a partly reusable launch system consisting of the reusable winged orbiter including three cryogenic main engines, an expendable external propellant tank for liquid hydrogen and liquid oxygen, and two solid propellant boost stages, the steel casings of which can be recovered for reuse. The orbiter provides space for a crew of up to seven astronauts and has a large payload bay that is sealed by two doors. Depending on the mission, laboratory units can be mounted in the payload bay. Today STS is solely used for building up and supplying the ISS. All commercial satellite transport missions were cancelled after the *Challenger* accident. Today's remaining three orbiters (*Endeavour*, *Atlantis* and *Discovery*) are aging, so STS is to be retired in the near future and replaced by a follow-on system.



Figure 3.1.13: *Atlas V (Source: NASA).*



Figure 3.1.14: *Space Transportation System with the Space Shuttle Endeavour (Source: NASA).*

3.1.5.2 Concepts for the Future

Reusability is considered to provide the major potential for future advancement of space transportation

Table 3.1.4: *Heavy-lift launch vehicles.*

	Ariane 5E/CA	Proton K/DM3	Atlas V 551	STS
Boost stages	2×EAP		5×SRB Block-B	2×SRB
Length/diameter	31 m/3 m		20.4 m/1.5 m	45.46 m/3.77 m
Take-off mass	273 000 kg		46 300 kg	590 000 kg
Propellant mass	240 000 kg		42 600 kg	502 000 kg
Propellants	HTPB-Al/AP		HTPB/AP	
Burning duration	140 s		n.n.	123.6 s
Main engine	1×P240		Solid motor	Solid motor
Thrust (vac/sl)	n.n./6360 kN		1655 kN	11 790 kN/n.n.
I_{sp} (vac/sl)	273 s/n.n.			267.3 s/n.n.
Remarks	Parachute recovery possible			Parachute system for recovery
First stage	1×EPC	1×liquid propellant stage	1×CCB	1×External Tank
Length/diameter	30.7 m/5.4 m	21.18 m/4.1 m	32.46 m/3.81 m	47 m/8.4 m
Take-off mass	185 500 kg	450 000 kg	305 566 kg	750 000 kg
Propellant mass	171 400 kg	419 410 kg	284 289 kg	721 000 kg
Propellants	LOX/LH_2	UDMH/NTO	LOX/kerosene	LOX/LH_2
Burning duration	540 s	120 s	250 s	
Main engine	1×Vulcain 2	6×RD-253	1×RD-180	
Thrust (vac/sl)	1350 kN/n.n.	1470 kN/n.n.	4152 kN/n.n.	
I_{sp} (vac/sl)	433 s/n.n.	316 s/285 s	337.8 s/311.3 s	
Remarks			Two-chamber engine	
Second stage	1 × ESC-A	1×liquid propellant stage	1×Centaur III	Space Shuttle Orbiter
Length/diameter	4.7 m/5.4 m	17.05 m/4.1 m	11.89 m/3.05 m	37.24 m/17.25 m
Take-off mass	19 200 kg	167 830 kg	22 760 kg	80 000 kg
Propellant mass	14 600 kg	156 113 kg	20 830 kg	14 200 kg
Propellants	LOX/LH_2	UDMH/NTO	LOX/LH_2	UDMH/NTO
Burning duration	970 s	210 s	900 s	522 s
Main engine	1×HM7b	4×RD-0210	1×RL-10A-4-1	3×SSME
Thrust (vacuum)	62.8 kN	582 kN	99.2 kN	2278 kN/1834 kN
I_{sp} (vacuum)	444 s	327 s	450.5 s	452.5 s/364 s
Remarks				Winged, reusable orbiter
Third stage		1×liquid propellant stage		
Length/diameter		4.11 m/4.1 m		
Take-off mass		50 747 kg		
Propellant mass		46 562 kg		
Propellants		UDMH/NTO		
Burning duration		230 s		
Main engine		1×RD-0210		
Thrust (vacuum)		582 kN		
I_{sp} (vacuum)		327 s		
Remarks		4 vernier RD-0214		
Fourth stage		Block D/DM		
Length/diameter		6.28 m/3.7 m		
Take-off mass		18 350 kg		
Propellant mass		15 200 kg		
Propellants		LOX/kerosene		
Burning duration		600 s		
Main engine		1×RD-58M		
Thrust (vacuum)		87 kN		
I_{sp} (vacuum)		353 s		
Remarks		7 ignitions		
Payload fairings				
Mass	2900 kg	n.n.	2255 kg	n.n.
Length/diameter	15.87 m/5.4 m	7.29 m/3.88 m	11.86 m/5.1 m	18.6 m/4.7 m
Payload	9000 kg GTO	4350 kg GTO 2600 kg GEO	n.n.	24 400 kg LEO 28°

systems. The reuse of the launch vehicle saves production costs; however, the complexity of the vehicle and the mission increase along with the necessary return capability. Typical objectives for reusable space launch systems are:

- Cost reduction
- Mission abort without loss of the launch vehicle and payload
- Return of payloads to the ground
- Higher reliability.

Higher reliability does not directly result from reuse, since the necessary return capability makes the launch vehicle more complex. However, a noticeable cost reduction can only be achieved with high reuse rates, which depend on very high system reliability. Apart from a high reuse rate, high reuse of the launch vehicle itself and/or its stages is necessary in order to achieve a noticeable cost reduction.

Because of the additional effort required to get the vehicle and/or stages back to Earth, reuse will be limited to launches into LEO (300–600 km level and 0–100° inclination), at least for the near future. Ongoing missions (GTO/GEO, Moon, interplanetary) will therefore continue to require expendable upper stages.

In analogy to aircraft a fully reusable single-stage space transportation system which delivers its payload together with an expendable upper stage into LEO is regarded as a goal. Various and sometimes very extensive US and European technology development and demonstration programs have shown, however, that substantial technological progress is still necessary before the target of a fully reusable single-stage vehicle can be achieved, particularly concerning:

- **Lightweight structures** (tanks and high-temperature thermal protection system)
- **Propulsion system performance** (rocket and air-breathing propulsion).

In the foreseeable future only multistage and partly reusable space transportation systems are feasible.

The Space Shuttle, in operation since 1981, was the first step toward a **reusable launch system**. It is a staged system and only partly reusable (the orbiter and casing of the solid propellant booster rockets). The Space Shuttle achieved neither cost reductions nor very high reliability. The return capability did, however, enable the retrieval of payloads. Presumably in 2010/2011, after approximately 30 years of operation, the Space Shuttle will be retired.

For future developments there are a multitude of options which lead to a variety of conceivable solutions (Figures 3.1.15 and 3.1.16). However, options are limited for influencing the substantial parameters affecting the design, namely:

- Partial or full reusability (for a majority of missions expendable upper stages are necessary)
- Number of stages
- Launch and landing method (horizontal, vertical, with/without propulsion, winged)
- Propulsion (rocket, air-breathing propulsion, combinations).

Based on experience with existing launch systems, extensive studies and technology activities, the following trends and limitations for future developments can be foreseen:

- Single-stage vehicles need new technologies.
- Reuse of boost stages leads to limited cost savings which do not justify the development and operating expenditure.
- Air-breathing propulsion is very complex and its integration into the overall design is demanding.
- Horizontal unpowered landing with wings is the feasible solution for the return of large rocket stages.

Besides these trends the substantial result of today's experience is the insight that to master the routine, reliable and cost-efficient operation of a reusable launch system, extensive **operational experience** is necessary, apart from theoretical understanding. Accordingly, further theoretical analysis and operational experience must progress in parallel. It is to be expected that several "generations" of launch systems will be necessary before operational skills of a "routine practice" level are achieved.

Theoretical analyses for future developments are summarized in **system concepts**. On the way toward optimization of concepts for future launch systems, numeric simulations play an important role. Today, very extensive and detailed analyses can be combined into an overall view of the system. NASA has

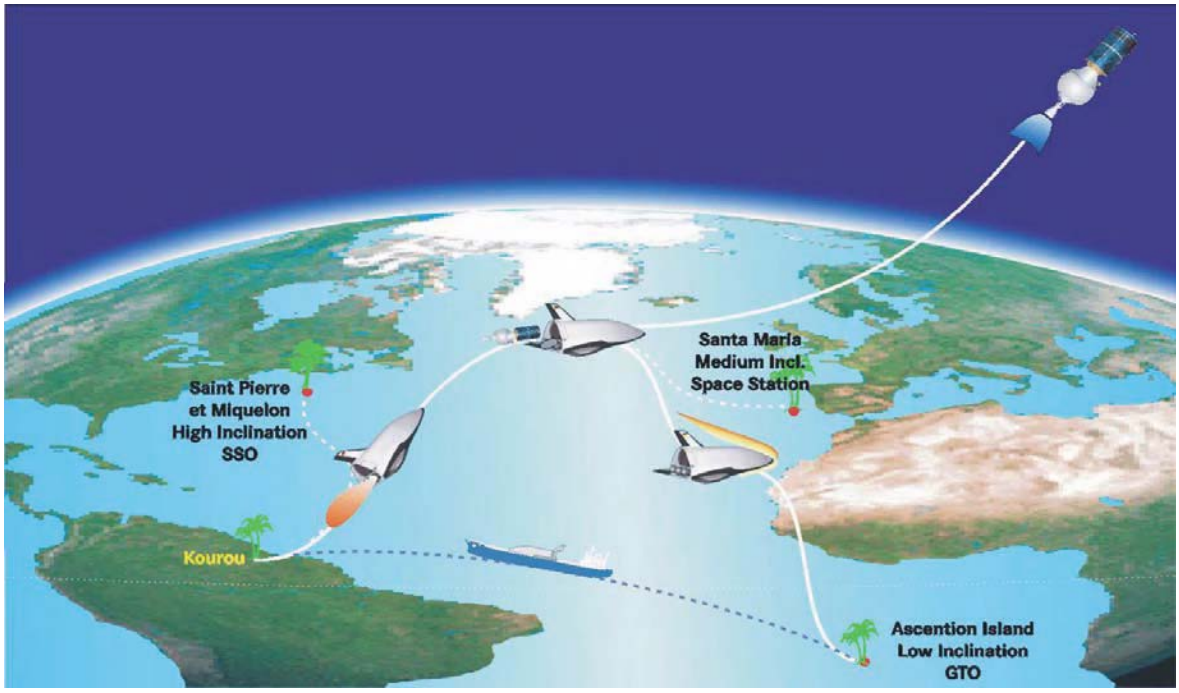


Figure 3.1.15: *The Hopper mission as an example of mission concepts for future partly reusable launch systems (Source: ASTRIUM).*

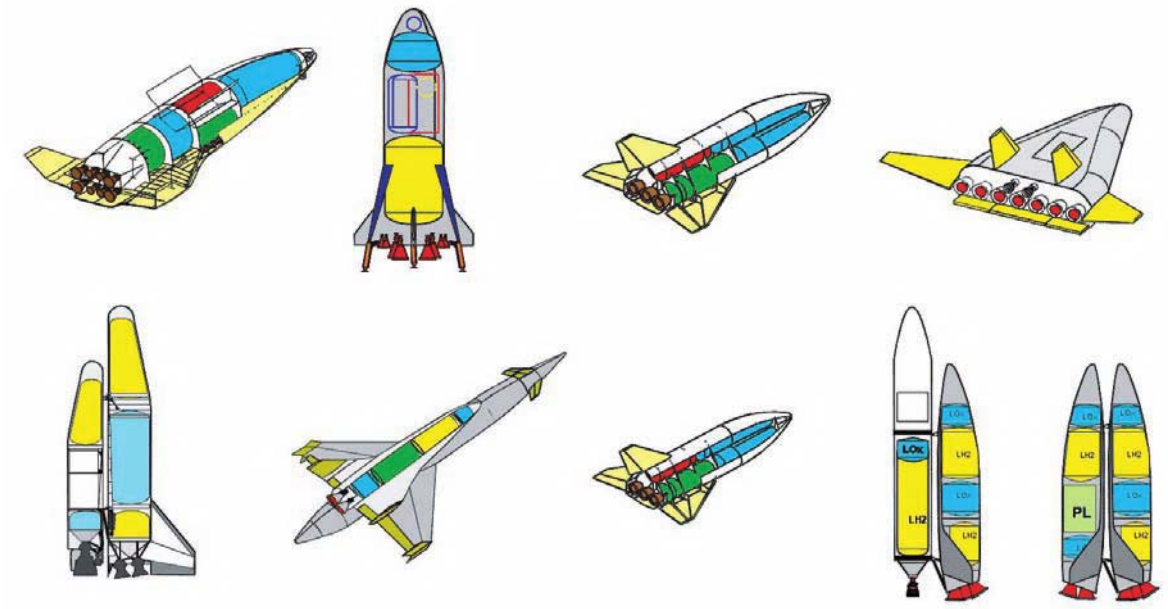


Figure 3.1.16: *Overview of substantial concept variants for future space launch systems (Source: ESA FESTIP study).*

implemented the interconnection of local **simulation programs** at different NASA centers via the Internet to be used locally (stand-alone) or as an integrated system. The attainable level of complexity is enormous and the number of modeled parameters reached (currently about 70 000) only tentatively indicates this complexity. The computed simulations are contributions toward improved understanding of possible future launch systems.

But already the definition of the optimization goal requires a commitment which cannot take into account all criteria of subsequent developmental decisions. There is the possibility to optimize performance by minimizing the take-off mass of the launch system for a given payload and target orbit. However, minimum take-off mass does not necessarily mean minimum costs, and minimum costs might be the costs for a single mission or those of the complete life cycle of the system, including development. In reality, however, not everything is developed from scratch. The use of available components saves development costs, although it can impose conditions on components or on the entire system leading to suboptimal solutions.

In a development, decision criteria which cannot be incorporated into a technical or cost simulation might also play a role. For a rational selection process, the definition of appropriate criteria and their weighting is essential. And weighting the decision criteria often leads to the real decision.

Bibliography

[3.1.1] Arianespace. *2006/2007 Market Survey Summary*, March 2007.

[3.1.2] EUROCONSULT. *World Market Survey of Satellites to be Built & Launched by 2016*, 2007 Edition.

[3.1.3] COMSTAC. *2007 Commercial Space Transportation Forecasts*, May 2007.

[3.1.4] Futron. State of the Satellite Industry Report, June 2006.

[3.1.5] Kuczera, H., Sacher, P., Dujarric, C. FESTIP System Activities – Overview and Status. *IAF-98-V-3.04*, October 1998.

[3.1.6] Obersteiner, M. Demonstration Demand and Logic towards Reusable Launch Vehicles in Europe. 12th European Aerospace Conference/3rd European Conference on Space Transportation Systems, November 1999.

[3.1.7] Spies, J. HOPPER – ein ASTRA Systemkonzept. *DGLR Jahrestagung 2000, DGLR-JT-2000-70*, September 2000.

[3.1.8] Obersteiner, M., Borriello, G. Future Launch Vehicles – Expendable and Reusable Elements. CEAS Conference on Materials for Aerospace Applications, December 2000.

[3.1.9] Brücker, H. Advanced Systems & Technologies for RLV Application – ASTRA. 2nd Symposium on Atmospheric Reentry Vehicles and Systems, Arcachon, March 2001.

[3.1.10] Spies, J., Grallert, H. Configurations Finding and Characterisation of ASTRA Reference Concepts. 2nd Symposium on Atmospheric Reentry Vehicles and Systems, Arcachon, March 2001.

[3.1.11] Obersteiner, M. Per PHOENIX ad ASTRA. Ausgewählte Systeme & Technologien für zukünftige Raumtransportsystem Anwendungen. *DGLR Jahrestagung 2001*, September 2001.

[3.1.12] Koelle, D.E. *Handbook of Cost Engineering for Space Transportation*, Rev. 2007. Ottobrunn: TCS, 2007.

[3.1.13] Isakowitz, S., Hopkins, J. *International Reference Guide to Launch Systems*, Fourth Edition. Reston, VA: AIAA, 2004.

[3.1.14] Huzel, D.K., Huang, D.H. *Modern Engineering for Design of Liquid-Propellant Rocket Engines*, Progress in Astronautics and Aeronautics, Vol. 147. Reston, VA: AIAA, 1992.

[3.1.15] Koelle, H.H. *et al. Handbook of Astronautical Engineering.* New York: McGraw-Hill, 1963.

[3.1.16] Fischer, H.M. *Europas Trägerrakete ARIANE.* Lemwerder: Stedinger Verlag, 2002.

3.2 Multistage Rocket Technologies

Jens Lassmann, Bernward Heese and Jörg Krüger

This section describes the available technologies and design criteria currently applied for rocket stages and their subsystems. Propulsion systems are only mentioned briefly since they are discussed in detail in Section 3.3. The interface between the feed system of a stage and its propulsion system is at the engine inlet valves. Stages powered by solid fuels will not be discussed here since there are only insignificant differences to the solid propellant propulsion described in Section 3.3.2.3.

3.2.1 Introduction and Overview

As previously mentioned, launcher systems typically consist of **several stages**. This approach makes it possible to optimize, in the sense of the Tsiolkovsky equation, the used chemical propulsion systems. After the burnout of each stage it is discarded and the engine of the next stage is ignited. This not only reduces the dry mass of the rocket, but also means that the design of each stage can be optimized for the requirements of the specific mission. An optimized staging determines the amount of fuel and the engine thrust for each individual stage, which has far-reaching consequences for the definition of the stage architecture, the propellant combination and the propulsion system. The distinguishing criteria and characteristics of staged launch systems are introduced below.

Location of Stages

Whereas the **lower stages** are optimized for thrust or power, the **kick stages** and **upper stages** are optimized for energy yield. For this reason, primarily cryogenic fuels are used in the latter (high exit velocity of the exhaust gas), whereas for lower stages solid fuels are more common (high impulse, high ejection mass flow, but low specific impulse). Kick stages only make sense if they either provide a significantly higher specific impulse than the apogee motor of the satellite, or can place it in orbit with more precision. **Transfer stages** for lunar or planetary missions are, in contrast to upper stages, far better adapted to the thermal conditions in orbit. In order to protect the super-insulation during ascent, transfer stages are usually located under the payload fairing.

Lower stages are longer and often have a larger diameter than rocket upper stages. The higher volume is needed to assure the high thrust required for ascent and the associated high propellants mass flow (see also Section 3.1).

Booster rockets can use either liquid or solid propellants. They are often operated in parallel with the lower stage. In the case of Ariane 4, high flexibility of the payload capacity could be achieved by using various combinations of liquid and solid fuel boosters. Both the US Atlas V and Delta IV launchers use so-called common core boosters (CCBs) in their heavy-load versions. Smaller solid fuel boosters (for Delta IV-M GEM-60 boosters) are replaced by liquid boosters of the same design as used for the main stage. This reduces production costs (learning curve, number of identical parts) as well as the acceleration load on the payload. However, such launchers reach their necessary final velocity only very slowly (long burn times for the lower stage and the booster) and the control of thrust is much more complex than when solid fuel boosters are used. For Delta IV-M rockets, so-called "cross-feeding" is utilized, whereby fuel is continuously pumped from the boosters to the primary stage during ascent.

Chemical Fuels

Stages powered by **liquid propellants** have a more complex design than in the case of solid fuels, since additional components and control equipment (such as valves, pumps and energy supply, measurement system) are necessary for the feed system, the fuel pressurization system and for the fueling operation. **Solid fuels** are used much more frequently in boosters and lower stages than in upper stages due to the limited flexibility and accuracy for orbit injection. After solid fuel stages have been ignited, there is normally no way to influence the firing duration (which is typically two minutes), since that was already determined by the geometry of the fuel when it was designed and fabricated. In the high-performance STS or Ariane 5 launchers currently in use, large **solid fuel boosters** are employed, which in the case of Ariane 5 provide up to 92% of the launch thrust (Figure 3.2.3). The Ariane 5 booster EAP, with its 37 tonnes of dry mass, height of 31 meters and load of 241 tonnes is one of the most powerful solid fuel boosters in the world. Efforts are being made to lower the launch costs of systems like Vega or Pegasus by extensive use of highly reliable solid fuel boosters.

Exotic propulsion concepts such as **solar thermal or electric propulsion** are only efficient in orbit and are accordingly only of interest for upper stages, kick stages or transfer stages. **Air-breathing concepts** require an oxidizer from the atmosphere during ascent and therefore can only be considered for lower stages.

Number of Ignitions

Most launcher systems are not using **reignitable stages** since refiring the main engines is problematic and

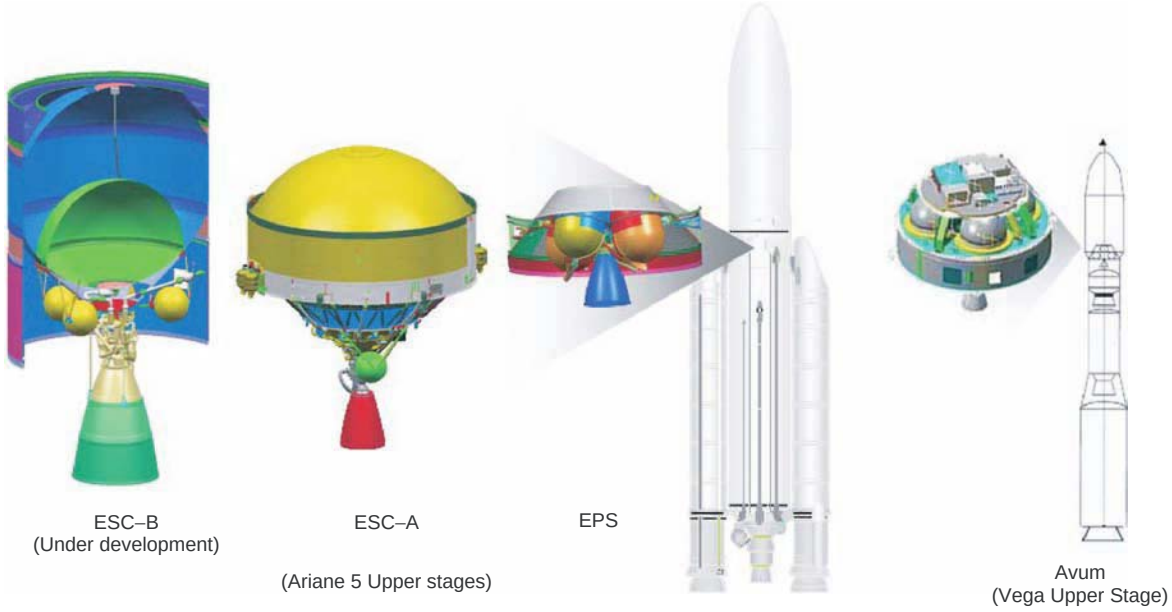


Figure 3.2.1: *Upper stages planned or in use for the European Ariane 5 launcher system and for Vega (Source: EADS).*

a source of failure. But the geographic position of the launch site is not always ideal for the desired satellite orbit, which requires that the upper stage be reignited because of the limited power of the launch system. Examples are the need to inject directly into a near-GEO orbit, or the descent of a stage to force its reentry into Earth's atmosphere. The Centaur upper stage of the Delta and Atlas launcher systems can be reignited, likewise the EPS-V upper stage of Ariane 5 and the Breeze and Block-DM upper stages of Russia's Proton rocket.

Reusability

Most launcher systems employed today have stages which cannot be reused, with the exception of the Space Shuttle orbiter. Fully reusable rockets do not exist at present.

Fuel Tank Architecture

One, but often several, fuel tanks are used for liquid systems (Figure 3.2.2). A distinction can be made between architectures in which all tanks are positioned in a central axis one above the other (1), or in which the tanks are toroidally and centrally positioned (2), or outcentered, circling the longitudinal axis of the rocket stage (3). Another question concerns which propellant to position closest to the engine. In order to locate the **center of gravity** toward the top, in the case of cryogenic propellants it is theoretically desirable to position the hydrogen tank closer to the engine in order to exploit the taller tank height of the hydrogen and the greater weight of the oxygen for effective thrust vector control. Nevertheless, for cryogenic upper stages the opposite is usually done (e.g., for the Centaur and ESC-A upper stages), since a heavy oxygen tank mounted below the hydrogen tank yields a smaller dimensioning loadcase and thus a lower mass for this stage.

For solid fuel stages and boosters the **engine housing** fulfills the two functions of fuel storage and combustion chamber. For large solid fuel boosters, as in Ariane EAP (Figure 3.2.3), the stage is produced in separate segments for manufacturing, transport and safety reasons and joined together only shortly before integration into the launcher .

Stage Autonomy

Whereas the US and European launcher systems usually position the on-board computer and avionics centrally in the upper stage of the launcher, Russian launchers often implement an autonomous approach

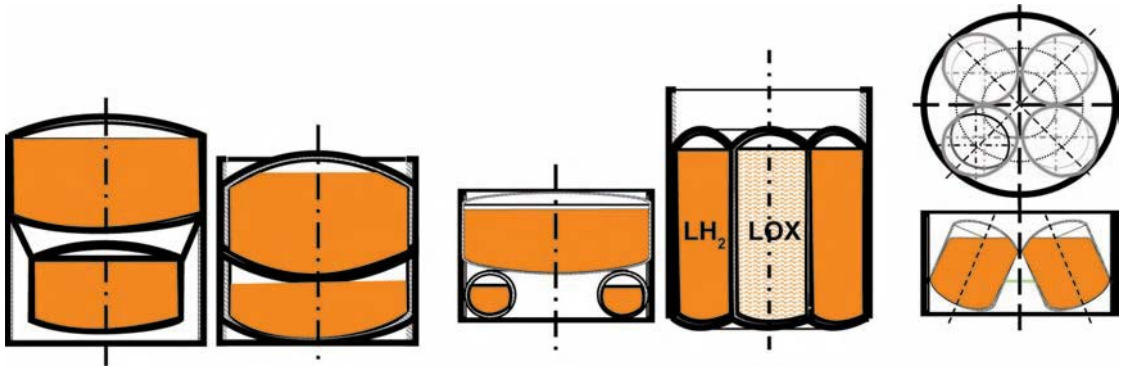


Figure 3.2.2: *Typical tank architectures for liquid fuel stages (Source: EADS).*

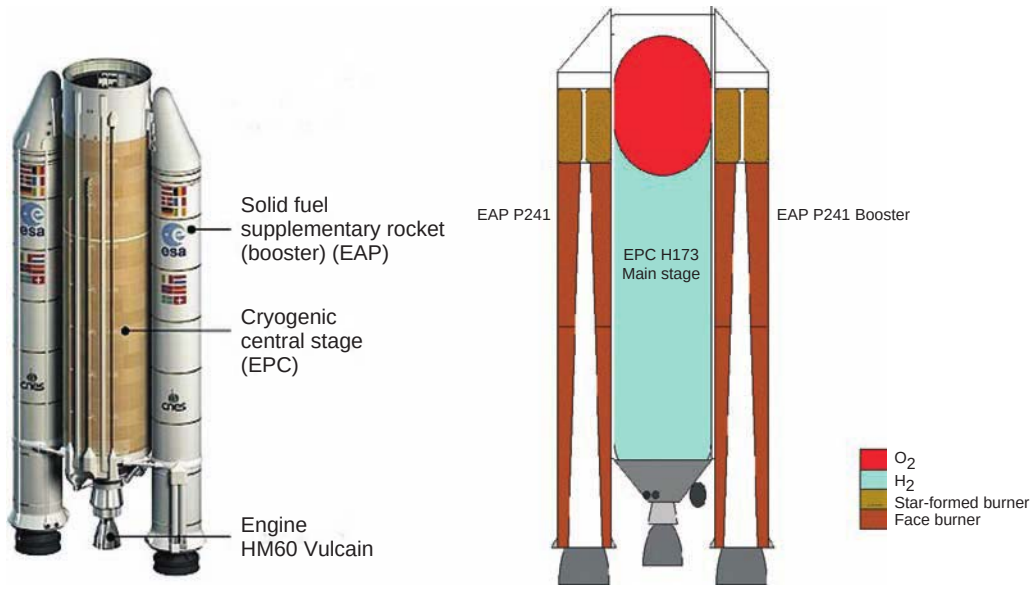


Figure 3.2.3: *Typical design of the booster and lower stages using the example of Ariane 5 (Source: ESA).*

with avionics in each individual stage. This autonomy can also be extended to the reaction control system (e.g., to correct for rolling) or to the energy supply (e.g., by using batteries). Only a low degree of autonomy has been realized in the Ariane launcher system in order to keep launcher mass low and systems reliability high.

Mass Index

Since propellants is constantly consumed during operation of the stage engines, the total mass of the rocket steadily decreases. If the thrust remains constant, then the acceleration of the rocket increases until engine burnout. According to the **Tsiolkovsky equation**, the increase in velocity of a rocket Δv is a function of the escape velocity and the mass ratio $M_0/(M_0 - M_P)$, with M_0 being the rocket mass at the time of ignition and M_P the propellant mass consumed between ignition and burnout. The more **propellant mass** a stage can transport with the same **dry mass**, the more the rocket can be accelerated, and the greater is the final velocity

at burnout of the stage. For this reason the dry mass of a stage should in principle be as low as possible. This value is accordingly taken as an indicator of the quality of performance optimization of a rocket stage. The technological effort required is always higher for an upper than for a lower stage, since the propellant mass is lower. Knowing the launcher's total mass is an important starting point when determining the performance of the launcher system. Table 3.2.1 provides information on the relative mass of stage subsystems for making rough mass calculations.

3.2.2 Mission Profiles and Operation

Typical Mission Profiles

The typical mission profile of a multistage rocket comprises the flight of the first stage(s), the separation from the first stage, the flight phases of the intermediate stage(s) (if available) and finally the flight phase of the upper stage. The upper stage mission is not finalized at shutdown, but completed only after the payloads have been successfully separated and the passivation maneuver is initialized.

A typical mission profile using the example of Ariane 5 is shown in Figure 3.2.4.

The current trend in new developments for upper stages requires reignition capability, which means that after the first shutdown further boost phases may follow, which are separated by ballistic flight phases. A mission profile of Ariane 5, which includes reignitions for the upper stage, is presented in Figure 3.2.5.

Prior to launch the propellant tanks of a launcher need to be loaded with liquid propellants and pressurized gases. In the case of storable propellants a launcher stage is loaded some days or even weeks prior to launch; in the case of cryogenic propellants loading starts on the launch pad and is continued until a few

Table 3.2.1: *Stage subsystem masses.*

Mass of stage subsystem as a percentage (%) of total dry mass[a]			
Subsystem	**Upper stage[e]**	**Lower stage[e]**	**Solid fuel stage[f]**
Structural index (dry mass to fuel load)	5.3 to 23	6.4 to 12	5.3 to 18
Engine and engine aggregates (incl. nozzles)	8 to 12	2 to 10	20 to 30
Fuel tanks (incl. insulation and fittings)	30 to 40[d]	40 to 60	55 to 85 (case)
He tanks for compression system	3 to 4	6 to 9	
Stage connections (lower interface)	10 to 17		
Stage connections (upper interface)	5 to 9		6 to 9
Engine thrust frame	0.6 to 2		
Avionics and fittings	3 to 5	0.3 to 1[c]	
Actuators for thrust vector control	0.5 to 1		
Ducts	1.5 to 2.5		
Attitude regulation system	0.4 to 1.8		
Other structures, cables and thermal insulation	10 to 30	5 to 20	2 to 4
Unusable propellant	0.5 to 2	2 to 10	
	Of loaded fuel mass[b]		
Power reserve	1 to 2.2	0.1 to 1.0	0

[a] Relative to dry mass on the ground, the stage dry mass before engine ignition is typically between 10 and 18% (structure index) of the propellant mass.
[b] The usable propellant mass minus what remains after engine burnout and margins for operation is about 92 to 98% of the propellant mass loaded on the ground.
[c] For an autonomous stage.
[d] Equipment for handling the propellant amounts to about 1 to 2% of the tank mass.
[e] Liquid: cryogenic or storable.
[f] Smaller value for upper stages with solid fuel propulsion; in this case the dry mass is almost linearly related to the load.

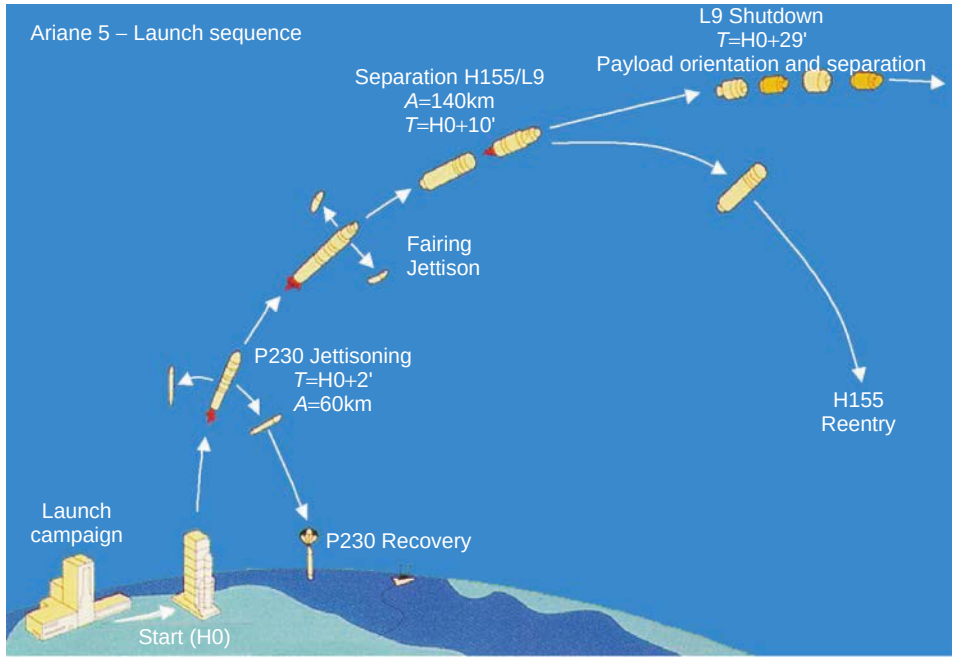


Figure 3.2.4: *Mission profile of Ariane 5 (Source: ESA).*

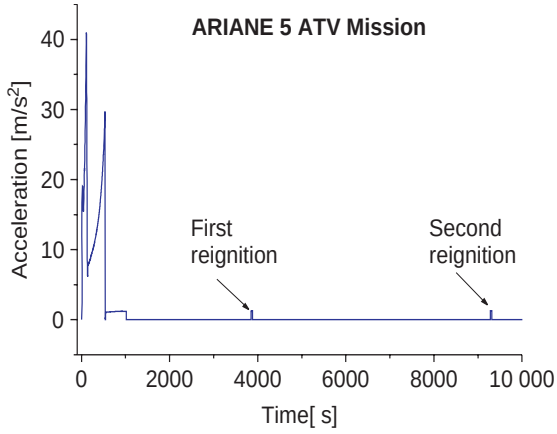


Figure 3.2.5: *Mission profile of Ariane 5 with reignitions of the EPS upper stage (Source: EADS).*

seconds prior to liftoff, in order to immediately replace vaporized propellants. In addition, the application of cryogenic propellants requires a chilling of the feed lines, the turbopump and the engine prior to ignition in order to avoid instantaneous vaporization of the propellants at the warm engine and an associated malfunction of the rocket engine.

The **ignition process** itself is monitored by computer and the rocket is cleared for liftoff when all engine parameters are within their nominal range.

During the **flight of the first stage** the later ignition of the following stage is already being prepared. If the following stage uses storable propellants, the tank pressure is adjusted to the correct level. If the following stage is operated with cryogenic propellants, relevant lines are chilled and the engine is initialized. The associated propellant vapors are dumped overboard by specialized valves.

The **stage separation** between two connected stages is a complex process which, besides the mechanical separation, also requires a distancing maneuver between the stages.

After successful release of the payload, the **passivation** of the last stage is initiated, which involves a dedicated pressure relief of the propellant and pressurant tanks. This maneuver is performed in order to avoid undesired explosion of the tanks of an upper stage when in orbit, in which the upper stage can remain for several months. A dedicated deorbiting maneuver avoids the risk of an explosion.

In case the launcher deviates from its nominal trajectory or reveals malfunctions in the propulsion system, the launcher is neutralized, in other words destroyed. Such a neutralization maneuver is, however, only initialized prior to reaching a stable orbit, since the particles and pieces would endanger other spacecraft as space debris.

Ballistic Phases and Reignitions

For the coming years a need for upper stages with **multiple reignition capability** is predicted. With reignitable upper stages satellites can be placed very close to their final orbit and the propellant stored in the satellite can be used for other purposes than for transfer or circularization maneuvers. Further applications for reignitions are resupply flights to the International Space Station (ISS), after which the upper stage needs to abandon the ISS orbit.

The capability to reignite a rocket upper stage requires that:

- The engine must be capable of reignition.
- The propellants must be at the right pressure and temperature for the engine.
- The stage must remain operational for a long time; energy supply and propellant losses due to vaporization must be taken into account.
- The stage must be able to navigate and balance disturbances in order to ensure maintenance of the required orbit (attitude control, avionics).

The capability to allow reignition is significantly determined by the engine and the applied **propellant combination**. In the case of nonhypergolic propellant combinations an additional ignition system for the combustion chamber and a potentially available gas generator are required.

For the ignition and during the ignition transient sufficient propellant needs to be available to supply the engine. If the liquid propellants are not retained within the lines and at the tank outlet, they will be arbitrarily distributed throughout the tank after a ballistic flight phase.

For correct reignition and operation of an engine, a sufficient quantity of the propellants needs to be available at the tank outlet (Figure 3.2.6).

Basically there are two possibilities **to position the propellants**:



Figure 3.2.6: *Liquid distribution within a propellant tank after shutdown of the main engine. (Source: EADS).*

- The propellant tank or the stage is able to collect a sufficient quantity of propellants at the tank outlet by itself; for example, by implementation of appropriate propellant management devices (PMDs) such as surface tension tanks.
- The propellant needs to be settled at the tank outlet by specific attitude maneuvers realized with small thrusters.

After successful ignition the upper stage is accelerated by the thrust of the engine, which also causes settling of the remaining propellants at the tank outlet.

Especially with a high thrust level at the main engines, considerable sloshing (**geyser effect**) at engine shutdown may occur due to the rapidly changing acceleration vector, which can result in heavy bubble formation within the propellant and, as a consequence, a propellant quality which is no longer suitable for the engine. Maneuvers for attitude control or payload separation act in a similar way; they are always associated with a turning of the stage or a counter impulse at the instant of payload separation. In all these cases the propellant is induced to slosh.

The **ballistic flight phase** which precedes reignition can have, depending on its duration, a significant impact on the conceptual design of an upper stage. Small, continuous consumption of all kinds, such as pressurants or electricity for navigation or data acquisition, requires additional storage capacity.

Depending on the duration of a ballistic flight phase, **heat loads** can also significantly impact the upper stage. Equipment exposed to solar illumination continuously warms up and requires additional insulation; on the other hand, equipment exposed to deep space continuously cools down and also requires additional insulation – in particular cases it even needs to be heated.

Heat loads also act directly on the propellants within the tanks. While the heat impact in the case of storable propellants is of less importance, it has fundamental importance in the case of cryogenic propellants. With these propellants heat loads cause an increase of the propellant temperature and corresponding steady evaporation. A possibility to further utilize heated propellant is the dedicated **depressurization maneuver** of the propellant tanks. During this maneuver the propellant temperature is forced to correspond to the pressure inside the tanks (relation of temperature versus pressure according to the saturation curve). However, boil-off losses occur during such maneuvers, which reduce the amount of fuel available for propulsion. The application of thermal insulation has a technical limit, which restricts the duration and number of ballistic flight phases with subsequent reignitions (especially for long-duration missions).

3.2.3 Components and Subsystems

The main subsystems of rocket stages are described below. An initial classification can be made following the categories in Table 3.2.2.

Propulsion System

The propulsion system is the most important subsystem of a rocket stage. It includes all those components which contribute to **thrust generation**. An overview of propulsion systems and engine cycles is given in table 3.2.3. The **number of engines** is a major design consideration affecting the functional requirements for the tank and the mechanical interface with the stage, which is the function of the so-called thrust frame. The turbopump inlet is the functional boundary between the engine and the stage propulsion systems. With more engines, less space is available for positioning auxiliary devices and supplementary tanks, for example for pressurized gas supply. Thrust vector control and the feed line routing become more complex. For the Atlas Centaur rocket, upper stages are designed with two motors (see Figure 3.2.7), but it is far more common to have designs with several engines in the lower stages,

Table 3.2.2: *Subsystems of rocket stages.*

Subsystem	Variants and distinguishing characteristics
Propulsion system	• Selected engine cycle, nozzle design for liquid propulsion • Number of engines
Feed system	• Pressurized gas • Pumps
Tank compression system	• Via inert gas (cold, warm, supercritical) • Vaporizing liquids • Vaporized propellants • Gas generators • Inert gas storage/production
Tank/propellant storage system	• Tank shape • Common or separate tank bulkheads • Reinforcements, ribbing, strengthening
Structures	• Interstage structures, intertank structures • Shroud and fairing • Payload adapter • Thrust framework, shell construction
Reaction control system	• Cold gas systems, hot gas systems • (Attitude control, fuel setting)
Avionics, energy supply	• Computers, telemetry, navigation • Batteries and accumulators
Stage separation system	• Separation cord • Pyrojacks • Hot gas systems
Thrust vector control and actuators	• With swiveling main engine • By thrust variation in one of several main engines (clustering) • With additional small thrusters • By influencing engine flow • With aerodynamic fins
Valve control system	• Pneumatic • Electric

Table 3.2.3: *Stage Propulsion Systems.*

Stage	Engine	Cycle	Propellants
Atlas V	RD-180	Staged Combustion	LOX/RP
Shuttle	SSME	Staged Combustion	LOX/LH2
Ariane 5 EPC	Vulcain 2	Gas Generator	LOX/LH2
Ariane 5 EPS	Aestus	Pressure Fed	MMH/NTO
Ariane 5 ESCA	HM7B	Gas Generator	LOX/LH2
Ariane 5 ESC-B (planning on)	Vinci	Expander Cycle	LOX/LH2

as in Ariane 4 (4 × Viking V), Zenit 3 (4 × RD-171) and Titan 4 (2 × Lr-87-11A). Launcher propulsion systems are described in detail in Section 3.3.

Propellant Feed System

For stages operated with liquid propellants there are two different ways in which the fuel can be conveyed from the tank to the engine (Figures 3.2.8 and 3.2.9) either pressure fed or pump fed. **Pressure fed systems** requires less equipment. But since tank wall thickness is directly proportional to tank pressure, large tanks (as required for cryogenic fuels, for example) imply considerable mass for the stage. For this reason, pressure fed systems are more common in small stages. For **pump fed systems** so-called turbopumps are used. The high-energy propellant now employed primarily in the upper stages are liquefied gases utilized at the saturation point of its vapor pressure curve. These liquids have a low boiling point and are extremely sensitive to heat loads introduced by the tank structure or feed lines. If the heat input upstream of the turbopump is too high, spontaneous vaporization may occur at the pump inlet. This so-called **cavitation** can lead to irreversible damage to the pump and must therefore be avoided. To prevent this effect, boost pumps are sometimes used inside the tank sump, to provide the necessary upstream pressure of the turbopumps, while at the same time keeping the pressure in the fuel tank at a low level.

Tank Pressurization System

The tank pressurization system supplies the propellant tanks with the amount of **pressurized gases** required to provide the needed upstream pressure for the engine propellant pumps (1–5 bar), or in the case of pressure fed systems feed, at the engine inlet pressure (10–30 bar). The stage tanks are pressurized to the minimal extent required for the engine either already on the ground (lower stages) or during ascent. They must be continuously fed with pressurized gas to replace the depleted propellant during operation of the stage motor in order to maintain the tank volume at constant pressure. Depending on the shape of the tank,

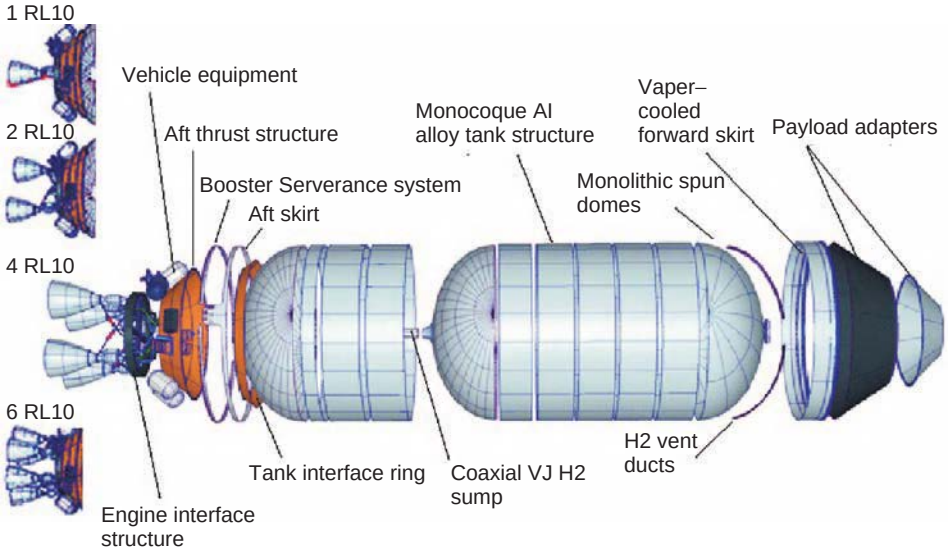


Figure 3.2.7: *Planned additions to Atlas Centaur for extended mission durations (Source: AIAA 2005-6738).*

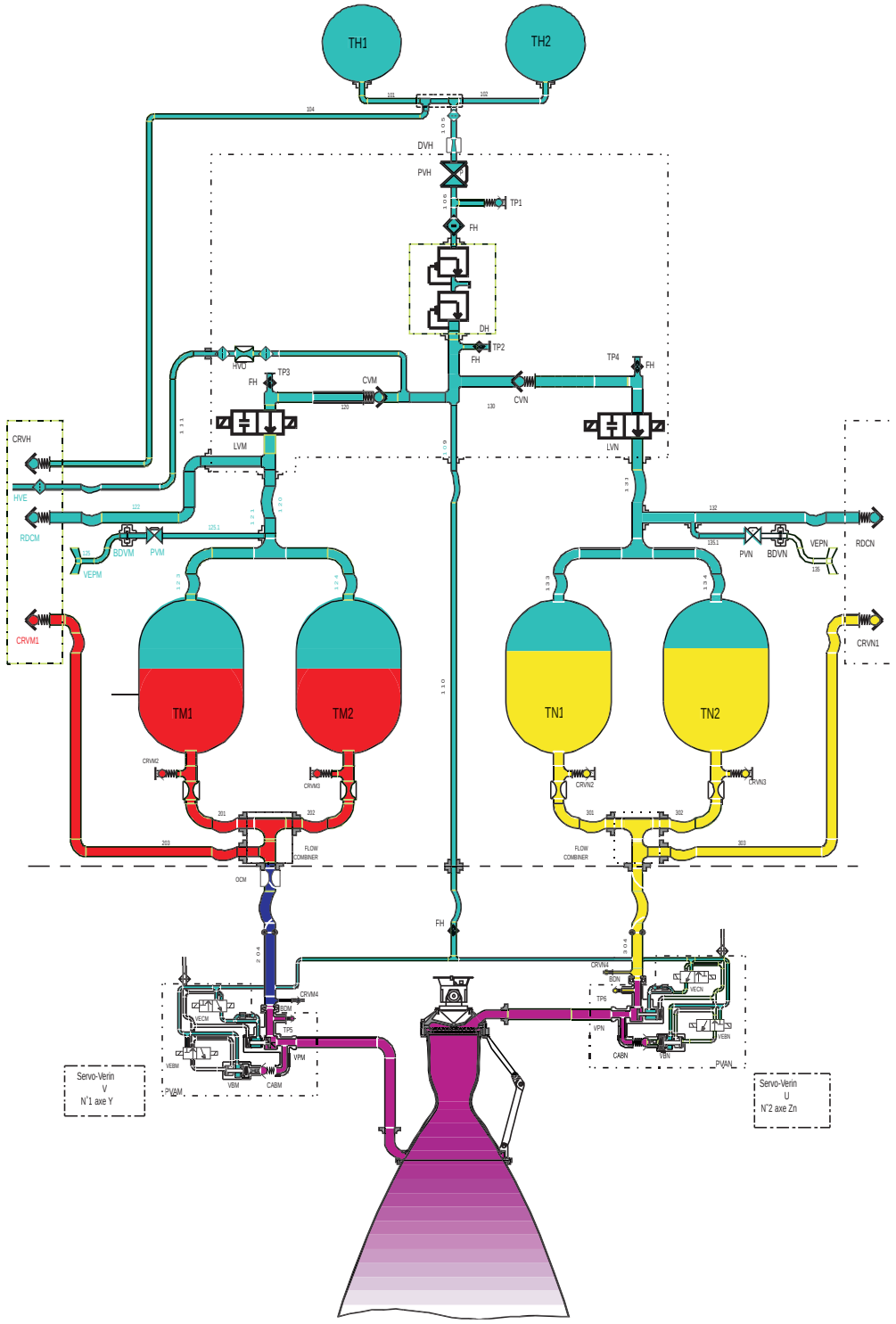


Figure 3.2.8: *Typical pressure fed propellant feed system using the example of the Ariane 5 EPS (Source: Arianespace)*

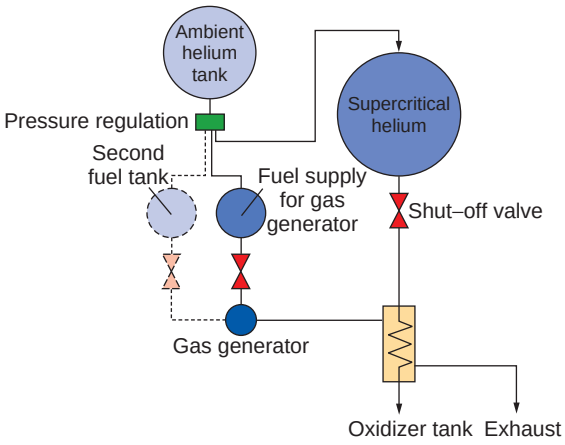


Figure 3.2.9: *Pressurized gas system with supercritical storage of helium.*

the initial gas volume in the fuel tank is typically 1–7% (upper stages) or 5–25% (lower stages and boosters) of the tank volume before the stage motor is ignited, depending on the transient pressure drop during the start phase. Inert gases, gaseous fuel combustion products or vaporized fuel are used as the **pressurization gas**. Vaporized fuel is obtained either from the main engine or from a separate gas generator, or induced in the ullage volume by regulated introduction of a hypergolic liquid. Inert gas has to be carried along in either gaseous or liquid form. Gaseous storage of helium in pressurized tanks (>200 bar) at 300 K ambient temperature or at low temperatures (77–150 K) is the usual choice. Helium is also used as a supercritical fluid at extremely low temperatures (4.2–10 K), for example for the lower stage EPC of Ariane 5. In general the pressurant gas should be introduced to the fuel tank at high enthalpy to keep the consumption low. For pump fed and cryogenic fuels, vaporized hydrogen is typically used for pressurization on the hydrogen side; helium is customary for the oxygen side. The third pressurization approach, using combustion products, is an option primarily for small, military rockets.

Propellant Tanks

The stage tanks store the propellant until it is burned in the engine to generate thrust. For solid fuel propulsion systems the tank is an integral component of the propulsion system. Because of the typically large fuel volume (especially in the case of cryogenic hydrogen, which has a density of about 70 kg/m^3), the fuel tanks are responsible for a large part of the stage's total structural mass (see also Table 3.2.1). For this reason they are usually **monocoque constructions** with thin walls. For cryogenic fuels, metallic materials are exclusively used at present, mostly aluminum alloys (titanium is incompatible with oxygen, and tanks reinforced with carbon fibers are still at an early stage of development). The tanks internal pressure helps to stabilize the walls against the usual bending and compression fluxes (balloon principle). The tank wall is usually load carrying regarding mechanical launcher loads; hence the tank walls and adjacent structures in the load path are strengthened with ribs or rings. When cryogenic fuels are stored, the tanks must be well insulated to keep evaporation losses low for gases which have been previously liquefied with considerable effort. In the case of a **common bulkhead**, either inner insulation or an evacuated tank bulkhead (as for the H10 upper stage of Ariane 4) can be used. Normally, cryogenic tanks are insulated externally with insulation foam with a low heat conduction coefficient and composite materials (MLI foil and fleece). Optimization is a challenge because of the loads in the launch and ascent phases. Operation times are limited by the unavoidable evaporation of the propellant. Besides lines for pressurization, fueling and supplying the engines, a number of devices are also located inside the tank. These are devices to position the fuel (PMD), suppress sloshing and swirling, and instruments to measure pressure, temperature, filling level, and possibly boost pumps to increase the feed pressure (Figure 3.2.10).

Having a common bulk head can help to reduce both stage height and dry mass. If cryogenic fuels are used, then because of the temperature difference between hydrogen and oxygen there must be suitable thermal insulation between the respective tank sections. For a pressure-stabilized common bulkhead, a permanent positive tank pressure differential must be assured from the concave to the convex side of the tank.

Structures

There are other structures integrated in the rocket stage in addition to the tanks for propellant and for the inert gases and additives which must be carried along. These include **interstage structures** connecting the stages, **intertank structures** between the various propellant tanks, housing for the stage avionics (in the **vehicle**

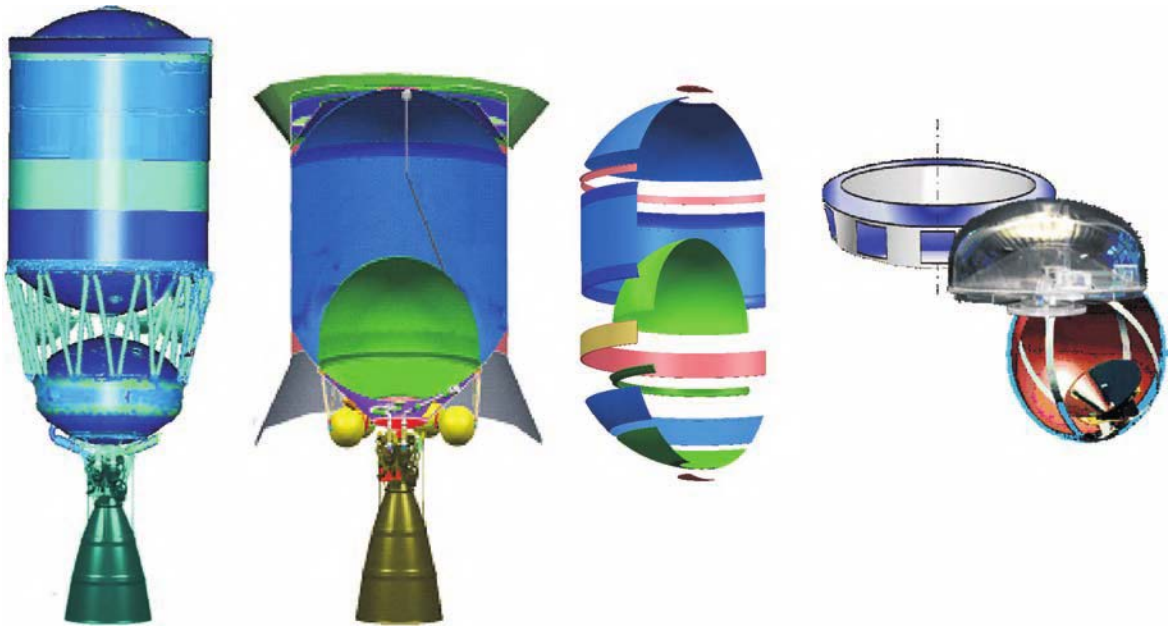


Figure 3.2.10: *Fuel tank with separate tanks, common tank floor and fuel management devices for positioning liquid propellant (Source: ASTRIUM).*

equipment bay), payload adapters, payload **fairing** and the **engine thrust frame**. These structures can be either monocoque or framework constructions. Either metallic materials or carbon fiber composites are used. Metal structures have the advantage that equipment, supports and lines can be easily attached; for fiber composites special inserts for screws and rivets must be provided. Fiber composites have the advantage of low heat conductivity. But if the structures are not too short or flat, then framework-supporting structures are preferred. If the rods are long, vibration dampers have to be included. If the installation angle is too high or the rods are too short, then the lateral loads must be absorbed in some manner, for example by using cross-beams or rings. Figure 3.2.11 shows typical design options using the example of two-engine thrust frames.

Reaction Control System

The functions of the reaction control system (RCS) using **small thrusters** range from attitude control to pre-acceleration, the settling of liquid propellant, to providing a velocity increment to separate stages and payloads, to carrying out maneuvers and to achieving small changes to the orbit (Table 3.2.4; see also Figures 3.2.13 to 3.2.14). There is more information on small thrusters in Section 4.4. A basic distinction is made between cold and hot gas systems. One characteristic is the use of stage fuel for attitude control, either by combüstion the fuel or by exploiting the pressure from the ullage of the fuel tank with the help of small thrusters. The latter approach is often found in cryogenic stages in order to make some use of the continuous and unavoidable evaporation losses.

Avionics and Energy Supply

The avionics and parts of the energy supply system are usually housed on a special platform, the **vehicle equipment bay**, and can be designed either centrally for the whole rocket or decentrally for each stage. Essential avionic components are the on-board computer; the flight control, telemetry and navigation systems; and the system for launch pad safety. The **flight control** system fulfills the functions of attitude regulation and flight stabilization, thrust vector



Thrust frame of the first stage of Zenit 1 with four RD-170 engines (Source: NPO Energomash)



Thrust frame of the Ariane ESC-A upper stage with HM7B engine (Source: ESA)

Figure 3.2.11: *Engine thrust frame concepts.*

Table 3.2.4: *Examples of reaction control subsystems used in upper stages.*

Launcher system upper stage	Main fuel	RCS fuels	RCS thrust [N]
Ariane 5 EPS, EPS-V	NTO/MMH	N_2H_4	6/8 × 400
Ariane 5 ESC-A	LOX/H_2	GH_2 Solid	6 × 58 4 × 3750
Atlas Centaur (V400, V500, Atlas IIIA/B)	LOX/H_2	N_2H_4	8 × 40 4 × 26.7
HII-A second stage	LOX/H_2	N_2H_4	6 × 50 2 × 50 (4 × 50) (2 × 4)
Proton M/Angara 3/5 Breeze M	NTO/UDMH	NTO/UDMH	4 × 396 12 × 13.3
Soyuz T, TM, TMA	NTO/UDMH	NTO/UDMH NTO/UDMH H_2O_2	14 × 137 12 × 24.5 6 × 67

control, navigation and guidance, and the initiation and control of vehicle operations (monitoring, propellant management, pyrotechnic command lines, health monitoring). Usually an inertial system (IMU, gyroscope) is employed for **positioning and attitude control.** The **telemetry system** is used to transmit digital and analog data, usually in the S-band range, to the ground stations and from there to the mission control center. For some launch systems relay satellites (TDRSS) are already being used for communications. Partial redundancy, and for safety-relevant components full redundancy, is often implemented. Modern launcher systems use GPS/GLONASS to determine and adjust flight position from the IMU. The **range safety** system protects the ground facilities and the vicinity of the launch pad. For this purpose certain of the launcher's operating parameters are continuously monitored and if there are any anomalies or if threshold values are exceeded, measures such as engine shutdown or even controlled destruction of the launcher are initiated.

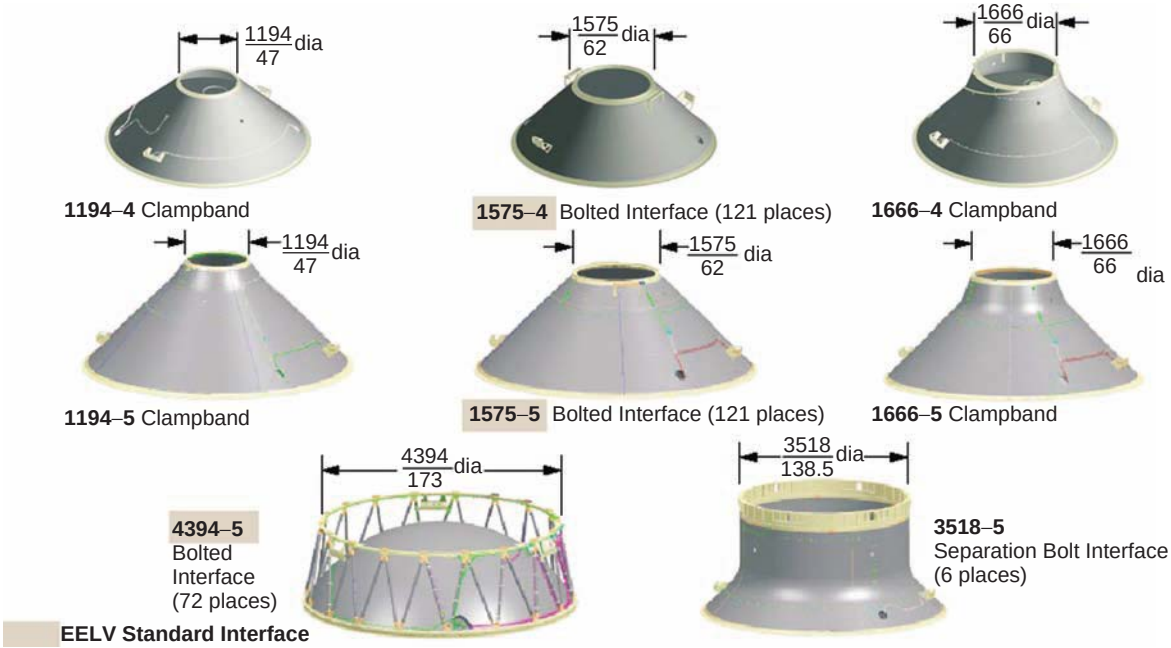


Figure 3.2.12: *Types of Delta IV payload adapters (Source: Boeing).*

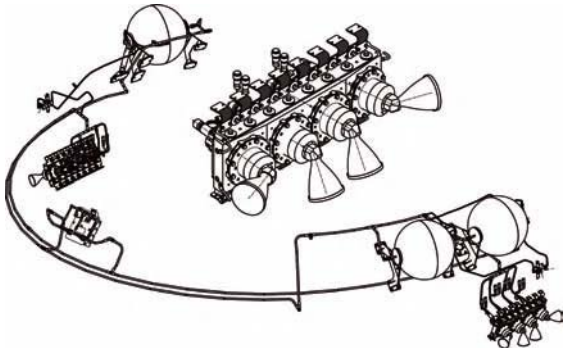


Figure 3.2.13: *Hot gas storage regulation system SCA used in the Ariane EPS (Source: ASTRIUM).*

The **energy supply** is via batteries which are charged immediately prior to launch. The Ariane 5 ECA launcher uses 60 V silver oxide–zinc batteries with an available current of 17 A for 65 minutes. There are also rechargeable nickel–cadmium batteries for the pyrotechnic equipment, the safety system, the telemetry and the power for the oil pumps which operate the actuators for the thrust vector control of the engines.

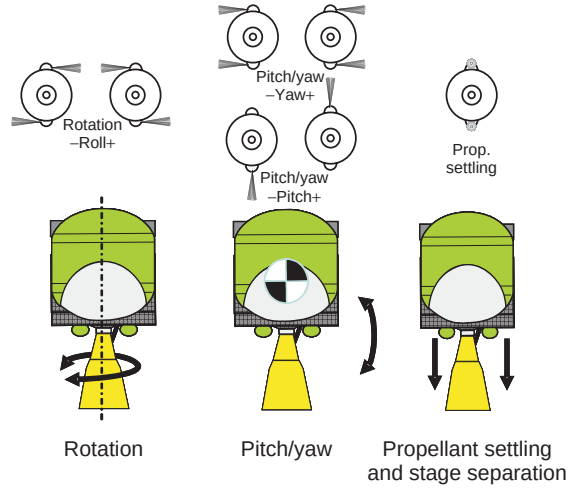


Figure 3.2.14: *Reaction control principle for upper stages (Source: ASTRIUM).*

Stage Separation Systems

The separation of rocket stages is usually initiated by **pyrotechnic systems.** A separation cord integrated in the interstage structure is ignited within milliseconds

for stage separation and at the same time the two stage segments are forced away from each other by spring force. In addition, small engines are often activated to accelerate the remaining launcher and separate it from the separated stage, or alternatively hinder the separated stage from continuing in the flight direction.

Thrust Vector Control and Actuators

The flight path of rockets has to be controlled and adjusted as long as they are in operation. Disturbances during the ascent, for example caused by wind, or by unbalanced centre of gravity position of the laŭncher, related to the pressure stagnation point and to the thrust vector have to be compensated. In rockets this is usually accomplished with the help of a **thrust vector control system**. Whereas aerodynamic stabilizer fin control (either in the air stream or in the engine jet stream) is only used today for small ballistic rockets. **Small engines** are effective for thrust control if they have a sufficiently high thrust level and efficient positioning in comparison to the main engine. Influencing the thrust vector by **injecting a liquid** into the diverging part of the nozzle is done in solid fuel boosters or stages. **Swiveling mechanisms** – usually Cardan suspension, a ball-bearing joint, or a notch – are used together with electromechanical or hydraulic **actuators** both for solid- and liquid-fueled stages. Their advantage over other designs is the very slight thrust reduction caused by the swiveling. The ESC-A upper stage of Ariane 5 uses a hydraulic–mechanical system, whereas in the Centaur launcher hydraulic actuators are used in the two-engine version and electromechanical ones in the one-engine version. If swiveling engines are used for liquid-fueled engines, then flexible or jointed lines have to be used for fuel feed in order to accommodate the swivel. Some Russian stages use the fuel to control the actuators instead of having a separate hydraulic fluid. In the case of several engines or large strap-on boosters, as in Ariane 5, these are frequently used to control the roll of the launcher during ascent.

Pneumatic Command System

The pneumatic control system provides the mass flow and pressure to operate valves. It is primarily used in cryogenic systems for safety reasons (reduced risk of explosion in the case of leakage). Helium is typically employed as a neutral inert gas.

3.2.4 Stage System Design Process and Technology

The **design of a rocket stage** has on the one hand to comply with all the requirements for the overall (launcher) system, such as main dimensions, performance and environmental and operating loads, and on the other hand all stage-internal subsystems such as propellant tanks, propulsions system, structures and control systems must be harmonized to form an optimized unit.

The stage contractor is accordingly responsible for the requirements and specifications of all equipment to be developed on the stage level, such as valves, lines and thermal insulation. The main subsystems are described in the following sections.

3.2.4.1 Design of the Overall Stage System

The main requirements for each stage, such as dry mass, engine thrust and the overall dimensions, are determined in the context of **launcher analysis** and forwarded to the stage contractors in the form of a preliminary specification.

These main requirements are the basis for the design of each stage. The **architecture of the propellant tanks** is critical, as it directly impacts the main dimensions of the stage and its dry mass. The volume sizing of the propellant tanks is determined by the propellant budget (Table 3.2.5). The mixture ratio of oxidizer and fuel depends on the choice of propellants and the type of engine. Consideration of the transient consumption for ignition and shutdown finally yield the total mass of loaded propellant. The final tank volume is determined by considering contingencies

Table 3.2.5: *Propellant budget of the Ariane 5 EPS.*

Contributions	N_2O_4 [kg]	MMH [kg]
Ignition transient	3.5	1.5
Usable propellant	6362.0	3105.0
Performance reserve	148.0	101.0
Shutdown transient	7.0	2.5
Residuals in tanks and lines	25.5	11.0
Loaded propellant mass	6546.0	3221.0

for the ullage volume, the volume for internal equipment and the performance reserve. In a next step the pressure distribution in the propulsion system is investigated and the layout for the pressurization system is chosen.

In most cases the **pressurization system** is combined with the pneumatic command system for valve activation. The activation of the pneumatic valves is initiated by the on-board computer. The amount of the pressurized gas mass is determined by the equivalence of mass within the "empty" propellant tanks at the end of the mission, complemented by contingencies for consumption of secondary order and losses.

The simplest way to supply the propellant tanks with pressurized gas is the application of a high-pressure vessel (often with helium storage) and depressurization to the required pressure levels of the propellant tanks via one or more pressure reducers.

This procedure has the disadvantage that the pressurized gas cools down during operation and the pressurization efficiency is reduced, which increases the amount of the pressurant mass. Alternatively, liquefied gases, such as hydrogen or helium, can be heated up at the operating (hot) engine and then supplied to the propellant tanks as the pressurant.

Concurrent to the design of the propulsion system, a **mass budget** (Table 3.2.6) for the overall stage is established, which comprises all elements of the complete stage. This mass budget is elaborated in a first step by simple hand calculations, data from catalogs and scaling equipment from already-built and similar rocket stages; in a later step during the development the preliminary mass budget is adjusted to the latest mass assessments from the hardware development. In case the requirements for dry mass cannot be fulfilled, the intended stage design is checked and the potential for mass savings identified. This process is iterative and can lead to a redesign of some subsystems or the complete stage.

Table 3.2.6: *Mass budget of the Ariane 5 EPS (dry mass).*

Subsystems	Mass [kg]
Structures	342.3
Propellant tanks	422.0
Propulsion system (incl. engine)	339.2
Lines	28.2
Electrical system	23.3
Pyrotechnical system	11.1
Supports and bolts	55.4
Thermal insulation	34.9
Total mass	1256.4

3.2.4.2 Impacts on the System Design

During the design of the stage systems several influences internally and externally to the stage have to be considered. Some of the important impacts are described in more detail below.

Loads

A distinction is made between mechanical and thermal loads; both types are further subdivided into environmental and operational loads. The determination of loads is performed individually according to the nature of their origin on either the stage or system level. Basically a stage must be designed to sustain all loads (including safety factors).

Mechanical Loads

The mechanical loads for the primary structures of each stage are determined on the launcher level, since the interactions of all stages need to be considered. The result of this coupled analysis is the requirements and inputs for the static and dynamic load assumption for the development of each stage.

External loads and self-induced and operation loads are combined at the stage level. This is done by superpositioning the external and internal loads of the stage, such as tank pressure loads and loads coming from operation of the engine, after which a dimensioning load case can be determined.

Thermal Loads

A distinction must be made between thermal environmental loads on the ground and those arising during the different flight phases. During the ground phase the launcher is exposed to the **thermal environment of the launch site**; the outer surface of the launcher is in direct contact with the external environment.

During the ascent phase the outer surface warms up due to air friction. If this heating up results in

unacceptable temperatures for the material of the structures, they need to be insulated. Special care is required for the areas of protuberances, as they significantly heat up during ascent.

During the active boost phase of the upper stage it is exposed to heat radiation coming from the hot nozzle (>1000 K) and the thermal conditions of the space environment. Depending on the architecture of the stage, the heat radiated from the engine nozzle is of importance, as all equipment near the nozzle will significantly warm up. In order to reduce this effect, a typical countermeasure is the implementation of a heat shield, which reduces the view area of the hot nozzle and thus the need for insulation.

During the payload separation phase, only loads from the space environment have an effect; these can lead to a temperature increase or decrease depending on the location.

Influences from Fluid Mechanics on Tanks and Lines

Fluid dynamic effects in the propellant tanks comprise mainly the thermodynamic conditions in the gaseous phase of the tanks, as well as sloshing effects and, specifically for cryogenic propellants, subcooling or warmup at the tank walls. For detailed dimensioning of these effects tests are unavoidable.

The **thermodynamic conditions** within the gas phase have via the ullage temperature a direct impact on the required amount of pressurized gas. For detailed determination of the required mass of pressurized gas, a thermodynamic model of the tanks is usually drawn up. The main parameters are the compression of the pressurized gas and the heat exchange within the gas phase with the propellant and tank walls.

Sloshing of propellants, initiated by internal or external loads, impacts the trajectory (disturbance forces) as well as the thermodynamic state within the ullage volume.

These mixing effects between propellant and ullage gas require tests for detailed assessment. Sloshing inside the tanks can be avoided or reduced by applying so-called antisloshing devices. Their geometry depends on the physical properties of the propellants and the required damping. An antisloshing device at the tank outlet of the Ariane 5 EPS is shown in Figure 3.2.15.

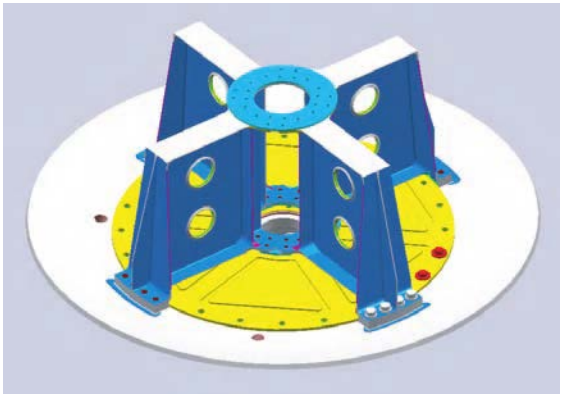


Figure 3.2.15: *Antisloshing device of the Ariane 5 EPS upper stage (Source: ASTRIUM).*

Fluid dynamic effects within the feed lines are less complex. In the case of storable propellants the application of one-dimensional calculation schemes is sufficient. In the case of cryogenic propellants a two-phase flow approach for transient behavior is required.

Influences from Aerodynamics and Trajectory

Aerodynamic effects, in a strict sense, are not a specific issue for a launcher, as the lift force is provided by the thrust of the engine and no other devices exist to create any lift forces. However, the **stagnation pressure** during the ascent phase is of importance, as drag forces directly impact the propulsion needed and affect the temperature of the external surfaces of the launcher. In order to protect the payload from stagnation pressure and temperature increase due to heating up, and in addition to reduce drag, the **nose of the rocket** (**fairing**) is aerodynamically shaped. However, this fairing is jettisoned as soon as the stagnation pressure is low enough to reduce the mass of the launcher.

Pollution/Cleanliness/Leakage

The term **pollution** summarizes contamination of all types coming from the launcher or the individual stages and released during ascent. The focus of these discussions is the contamination of the payload. Accordingly, requirements concerning contamination are driven by the payload. Contamination of the payload, for example the solar panels, has to be

prevented in order to avoid degradation of performance. Pollution of the payload can occur only after the payload protection (fairing) has been jettisoned, and when there is no axial acceleration. The sources of contamination for the payload are small particles and also molecules released from materials of the upper stage by outgassing. Furthermore, the operation of the RCS could be a source of potential contamination.

By **"cleanliness"** the internal pollution level of the stage is meant. The correct operation of the propulsion system of a launcher stage can be ensured only if the internal cleanliness level of all components, such as tanks, lines, valves and engine, satisfies the requirements. The required cleanliness level is derived from what is required to ensure the correct operation of valves and engines. This cleanliness level must then be ensured throughout the entire pressurization and propulsion system. This means that lines, equipment and also tanks need to be specifically cleaned and their cleanliness verified. The cleanliness requirements are verified as the allowable amount of particles within a defined volume of the cleaning fluid. Typical cleanliness requirements for lines of the propulsion system of a launcher stage are summarized in Table 3.2.7.

The use of **filters** within the propulsion system is avoided in most cases in order to avoid the risk of partial or complete clogging resulting in an undefined increase in hydraulic pressure losses.

"Leakage" describes the release of pressurized gases and propellants from lines and line connections. The impact of a leak can vary: leakage in the pressurization system corresponds to an unintended loss, while leakage in the propellant circuits can result in explosive mixtures within closed volumes.

For unavoidable leakage a specific "leakage budget" has to be established. Measures to minimize leakages are the application of appropriate connection elements and a tightness verification during acceptance testing of components and line circuits. Typical tightness requirements during acceptance testing of connection elements by application of helium are in the range of 10^{-5} cm^3/s.

Table 3.2.7: *Allowed particles in 100 ml of cleaning fluid.*

Particle size [µm]	Amount
≤25	No limit
26–50	≤15
51–100	≤5
101–200	≤1
>201	0

3.2.4.3 Subsystem Design

Thermal Design

The objective of the thermal design of launcher stages is to ensure acceptable temperature conditions for all equipment during all ground and flight phases. This analysis starts with the identification of all heat sources and sinks and in a next step, if possible, with an appropriate positioning of all equipment.

After the stage configuration is sufficiently advanced, a **thermal mathematical model** is established, which comprises all main components and structures. Thermal environmental conditions are then used to calculate the temperature distributions as a function of time for all components. The evaluation of these temperature distributions regarding the minimum and maximum temperature identifies the need for countermeasures. In case components already exceed their allowed temperature range during the ground phase, a conditioning by flushing with warm or cold gas, in most cases nitrogen or helium, is applied. These flushing gases are supplied from the ground facilities at the launch area and are no longer available after liftoff. During the flight phase only the application of insulation or capacitive cooling elements remains feasible. The application of electrical heaters is limited because of their additional electric power need.

Thermofunctional Design

The thermofunctional design of tanks and lines containing propellant is specific for **cryogenic stages**. The focus of these considerations is the warming up of the propellants due to heat loads which result in an increase of the vapor pressure and a decrease of the propellant density. The latter causes a mixture ratio shift of the engine, which can result in a reduction of the specific impulse and an increase of the temperature inside the combustion chamber. While the reduction of the specific impulse affects only the stage performance, the increase of combustion temperature can result in damage to the combustion

chamber. In the case of cryogenic propellants the thermal residuals in the tanks need to be minimized. Design aspects which reduce heat input are the stage architecture, the shape of the tanks and thermal insulation on the outer wall, in specific cases also on the inner wall of the tanks.

Geometrical Design and Layout
After finalization of the staging of the launcher, positioning of the control units and the tank configurations and connecting structures, the geometrical design of each launcher stage is started. The accommodation of **stage equipment** is accompanied by comparative studies (trade-offs). However, the accommodation of stage equipment remains an iterative process, as counteractions between the positioning of equipment can only be identified in the total layout.

The **design process** of a component or subsystem starts with layout schematics, which are then reviewed with respect to requirements and mass. When all requirements are met a design definition drawing of the concerned area is elaborated which comprises all essential details. As a last step interface drawings, manufacturing drawings and integration drawings are made. The described procedure is similar for each subsystem. If a component cannot be defined because its development is still underway, the design work is started using an envelope volume representative of the main dimensions and interfaces. According to the progress of the development of a component, this envelope volume is replaced with evermore precise definitions of the component.

In order to keep the later production costs low, simple and low-cost design (design to budget) needs to be considered from the start of the development. Also, the later integration process needs to be considered from the beginning in order to ensure a final integration process free of disturbances. A definition drawing of the EPS upper stage of the Ariane 5 is shown in Figure 3.2.16, while Figure 3.2.17 shows the final product after finalization of the development.

Mechanical Design
After finalization of the definition of the **main structural elements**, they are designed using mathematical models together with the **load requirements** from the launcher analysis and loads coming from the

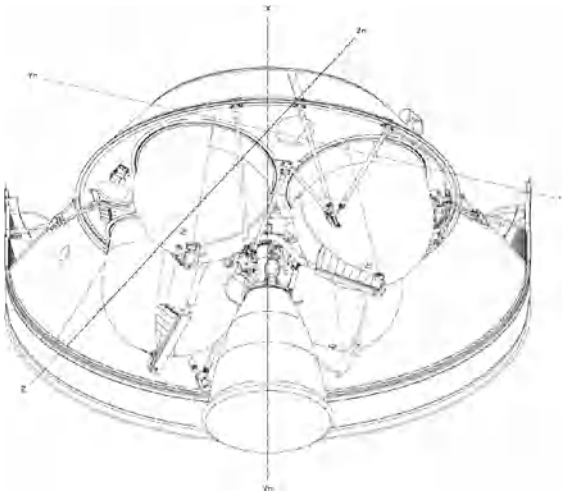


Figure 3.2.16: *Definition drawing of the Ariane 5 EPS (Source: EADS).*

operation of the stage. The result of the design process is the geometrical definition and the definition of the required thickness of the materials. The **choice of the applied material** is in most cases determined by manufacturing processes or predefined costs. After the design of all main structural elements, static and later dynamic mathematical models of these items are elaborated. By means of the static model and under consideration of the loads, the safety factors against overload are determined. With the help of the dynamic model and the dynamic excitations at the stage interfaces, the mechanical responses at all necessary stage internal locations are determined. The result of these calculations yields the local loads at particular locations, such as between lines and components, which are then used for further local analysis and design.

During the mechanical design process of a launcher stage other loads also need to be considered which are caused by **shocks, vibrations and acoustic noise**. These loads are sometimes challenging and, especially for sensitive components, can require the application of specifically developed damping elements.

Design of the Flight Control System
The flight control is in most cases performed by a **central control unit** (on-board computer) for the complete launcher. The control unit operates

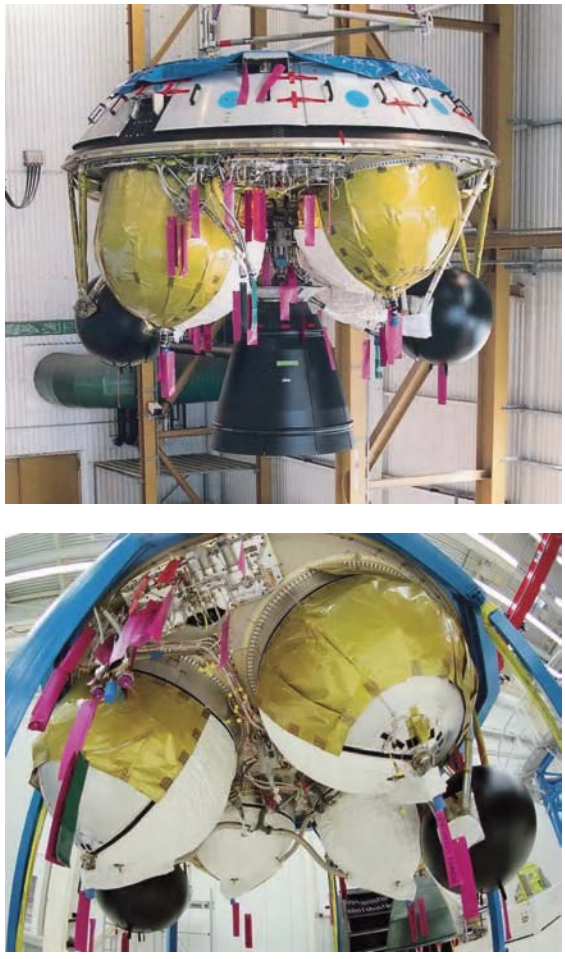


Figure 3.2.17: *The Ariane 5 EPS after development (Source: ESA).*

typically by means of a "target–performance comparison" of the predefined trajectory and corrects deviations caused by disturbances by applying control elements. The most important control elements are the thrust vector control unit and the attitude control system (see Section 3.2.3). The requirements for these control elements are determined on the launcher system level and comprise data about the thrust level, minimum thrust impulse as well as the total required impulse.

The control system continuously acquires information from each stage which allows assessment of the correct operating conditions. The acquired information includes the launcher acceleration, the engine chamber pressure, the filling level of the propellants inside the tanks and the trajectory. In case the monitored parameters are outside their predicted range, a premature shutdown of the concerned stage can be initiated.

Stage Optimization and Performance Determination

The performance of a launcher is determined by the **payload mass** which can be injected into a predefined orbit. All relevant data such as dry mass, center of gravity, characteristic of the thrust profile, etc., for all stages is then processed in a **trajectory program** and all parameters of the trajectory are then determined. In case the required payload mass cannot be reached, measures to increase the payload capability are sought, such as reduction of the dry mass of the stages and optimization of such operating parameters as thrust level, mixture ratio, propellant residual masses or specific impulse. The objective of the optimization process is to minimize the mass of each stage at shutdown. Besides the dry mass of each stage, the propellant residual masses have a large impact. This concerns especially cryogenic stages which form thermal residuals caused by heat entry.

3.3 Propulsion Systems

Dirk Rüdeger Albat, Oskar Josef Haidn and Günter Langel

Propulsion systems for rocket engines are discriminated according to their propellant combinations: **solid**, **liquid** and, as a combination of both, **hybrid propulsion systems**. Liquid propulsion systems are subdivided into monopropellant and bipropellant systems. The first are designed as cold gas systems or apply catalysts to decompose the fuel. Examples of typical monopropellant systems are hydrazine (N_2H_4) or nitrous oxide (N_2O). Bipropellant systems have a broad variety of oxidizers and fuels, see Section 3.3.1.2. Depending on mission requirement and staging of the launcher, the different propulsion systems can be

classified into four applications: **booster**, **main and upper stage** engines, **satellite and attitude control systems**.

Typical rocket engines are combustion devices whose energy release exceeds by far what is materialized in conventional fossil-fired or nuclear power plants. While the thermal power of solid fuel rocket engines may exceed 30 GW, those of liquid propellants very seldom reach 20 GW and most of them range between 3 and 10 GW.

3.3.1 Chemical Propulsion Basics

Any chemical propulsion system is based on energy released inside a combustion chamber. The resulting high pressures and temperatures of the reactants the desired thrust in the chamber and produce nozzle (see Figure 3.3.1).



Figure 3.3.1: *Vulcain 2 engine at P5 test facility in Lampoldshausen (Source: ESA).*

3.3.1.1 Basic Equation

The basic relationship, called the **rocket equation**, was first derived in 1903 by Konstantin Tsiolkovsky and describes the principles of rocket propulsion. One-dimensional flow, see Figure 3.3.2, yields the thrust of such a system:

$$F = \dot{m}\, u_a + (p_a - p_\infty)\, A_a \ [\mathrm{N}] \tag{3.3.1}$$

with u_a and p_a the velocity and the pressure of the exhaust gases at the nozzle exit A_a, and p_∞ the ambient pressure. The first term in Equation 3.3.1 represents the **momentum part** and the second term the **pressure part** of the thrust F.

$$F = \dot{m}\, c \ \text{ oder } \ F = \dot{m}\, C_F\, c^* \tag{3.3.2}$$

where:

F = the thrust,
C_F = the thrust coefficient,
c = the effective exit velocity,
c^* = the characteristic velocity.

The frequently used term **specific impulse**, I_{sp}, can be computed according to

$$I_{sp} = c / g_0 \ [\mathrm{s}] \tag{3.3.3}$$

where g_0 is the gravitational constant. A more detailed look at the characteristic **velocity** c^* and the **thrust coefficient** C_F reveals the effect of fluid properties and operating conditions on the thrust relation (Equation 3.3.2). The characteristic velocity c^* relates the combustion chamber pressure to the amounts of propellants burnt and thus reflects the fuel energy and the combustion efficiency. Equation 3.3.4 shows that c^* increases with the gas temperature T_b in the thrust chamber but decreases with the

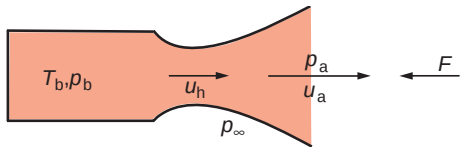


Figure 3.3.2: *Principle of an ideal rocket.*

Table 3.3.1: *Typical values of some characteristic coefficients.*

T_b [K]	p_b [MPa]	M [kmol/kg]	c^* [m/s]
1000–3800	1–26	2–30	900–2500
ε [–]	k [–]	C_F [–]	I_{sp} [s]
15–280	1.1–1.6	1.3–2.9	150–480

molecular weight M and isentropic coefficient k of the exhaust gases:

$$c^{\star} = \left[\frac{1}{k} \left(\frac{k+1}{2} \right)^{\frac{k+1}{k-1}} \frac{R\, T_b}{M} \right]^{1/2} \tag{3.3.4}$$

$$C_F = \left\{ \frac{2\,k^2}{k-1} \left(\frac{2}{k+1} \right) \left[1 - \left(\frac{p_a}{p_b} \right)^{\frac{k-1}{k}} \right] \right\}^{1/2} + \left(\frac{p_a}{p_b} - \frac{p_\infty}{p_b} \right) \varepsilon \tag{3.3.5}$$

with ε being the ratio of the nozzle throat area to exit area and R the universal gas constant. The thrust coefficient C_F increases with decreasing expansion coefficient k. Both the ratio of chamber pressure to nozzle exit pressure and the difference between exit pressure and ambient pressure are important. However, although high exit velocities are favorable, any expansion below the ambient pressure lowers the thrust coefficient. Table 3.3.1 summarizes the range of variation of characteristic coefficients for rocket engines.

3.3.1.2 Propellants

The equations and the arguments about the characteristic velocity and thrust coefficient mentioned previously allow for a **propellant classification**. The goal is to achieve high combustion chamber temperatures, and low molecular weight of the reactants. Furthermore, the propellant choice might depend on its density or temperature, the engine cycle, or on certain mission requirements. While the **combination of H_2/O_2** is generally considered the best choice for upper stage engines, kerosene/LOX (LOX = liquid O_2) or solid propellants are frequently preferred for booster engines, which operate primarily within the atmosphere.

Although from a performance point of view a broad variety of propellant combinations are feasible, only the most frequently used are presented here; for more information see [3.3.1]. Hypergolic combinations of N_2O_4 with different hydrazine compoŭnd and derivatives have been frequently used for a long time because of their superior storability despite their minor performance values, serious toxicity issues and related handling problems. In the meantime LOX/kerosene has become the combination of choice for large liquid propellant booster engines; the most impressive engines are the F1 engine of the Saturn V and members of the RD-170 engine family.

The influence of **mixture ratio** and **propellant combination** on the **specific impulse** of an engine is depicted in Figure 3.3.3 and allows classification of the propellants. The cryogenic propellant pair of hydrogen and oxygen exceeds by far (30%) any other combination. Oxygen and simple hydrocarbons such as methane, propane or kerosene build a group of propellants with similar performance. Within this group the specific impulse decreases with an increasing nŭmber of carbon atoms. Pairings of oxygen and simple alcohols build a second group with performance values similar to the hypergolic mixtures, such as N_2O_4 and MMH, with about a 10% performance decline. For comparison Figure 3.3.3 also shows the performance of kerosene and ethanol with 90% hydrogen peroxide (H_2O_2).

3.3.2 Types of Engines

Rocket engine cycles are mainly divided into three categories according to their fuel feeding technologies, see Figure 3.3.4. Engines requiring lower power levels are mainly fed using tank pressurization via gas stored in high-pressure tanks, while most of the engines used in launchers work with turbomachinery to provide elevated combustion chamber pressure.

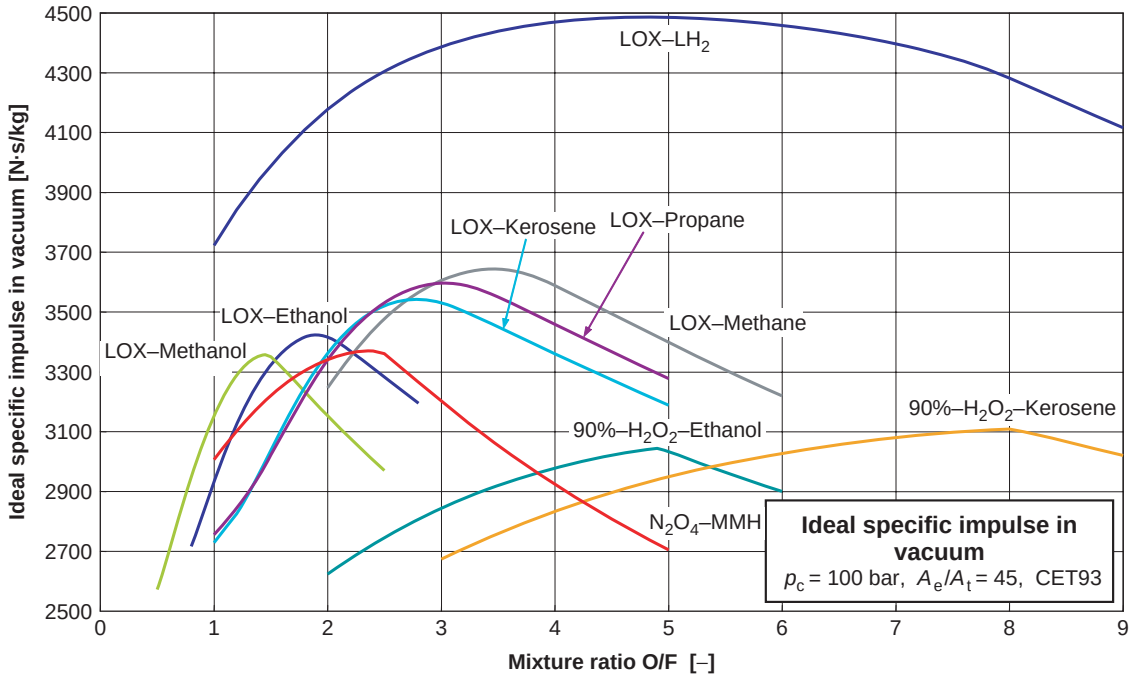


Figure 3.3.3: *Specific impulse of various propellant combinations.*

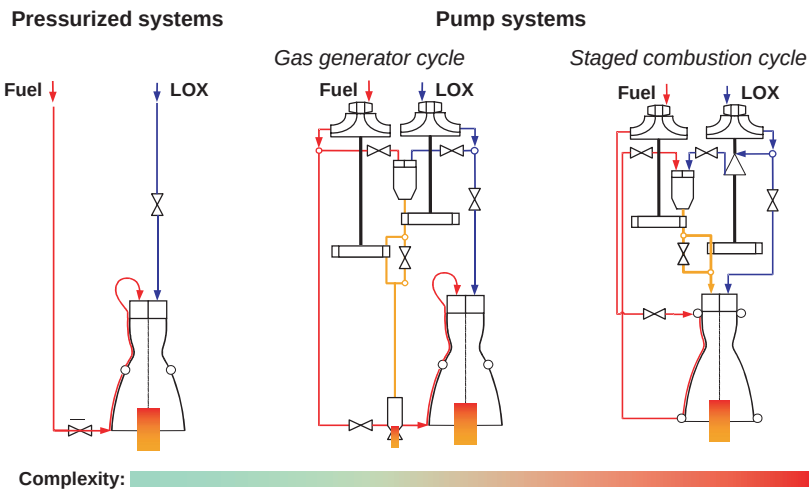


Figure 3.3.4: *Engine systems.*

The influences of combustion pressure, nozzle expansion ratio and the specific impulse are important for performance characterization. The latter are the main drivers of engine performance, while combustion pressure sizes the dimensions of the engine, see Figure 3.3.5.

3.3.2.1 Pressure-Fed Systems

Pressure-fed engine systems can have either **internal** or **external pressurization**. The first option depends on evaporation of a propellant by the addition of heat. This is often applied in monopropellant systems.

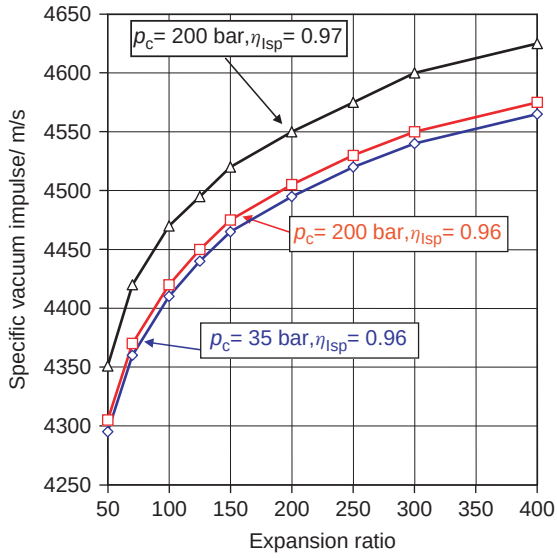


Figure 3.3.5: *Specific impulse as a function of the nozzle expansion ratio ε (pressure and efficiency are parameters).*

The latter option uses high-pressure gas (helium) stored in pressure bottles. These systems are limited by the propellant tank technology and are almost exclusively used for upper stage systems with storable fuel.

3.3.2.2 Pump systems

Table 3.3.2 compares various engine versions. The **expander engine** cycle is a specialty because the energy for the pumps is not produced by means of a gas generator, but is delivered by one of the propellants itself, which is heated in the cooling channels of the combustion chamber. For **pump-fed engines** the principle of turbochargers is usually applied. Hot gas is produced in a gas generator and expanded in a turbine to operate the pumps to increase the propellant pressure. If after expansion the gas is led into the main combustion chamber the cycle is called staged combustion; if the hot gas is fed through external exhaust pipes the system is called a gas generator cycle. There are two classes of gas generators: LOX-rich or fuel-rich. The majority of gas generators are operated fuel-rich. Only Russian engine systems operating with the propellant combination of LOX/kerosene use the oxidizer-rich version, and there is only one system known worldwide which applies both options,

Table 3.3.2: *Comparison of engine cycles.*

	Open cycle rocket engine (gas generator cycle)	Closed cycle rocket engine (staged combustion cycle)	Expander cycle rocket engine
Advantages	• Simple interface condition between thrust chamber and turbopumps • Independent development of the subsystems, turbopumps and thrust chamber • Moderate pump outlet pressure	• Highest combustion chamber pressures without performance loss • Compact turbine design • No separate turbine outlet	• Simple engine design due to lack of a preburner • Ignition of main combustion chamber only • Low turbine inlet temperatures
Disadvantages	• Engine system performance loss driven by fuel-rich gas generator operation • Low turbine efficiency • Limits on maximum combustion chamber pressure due to increase of secondary flow rate	• Complex development effort due to close interaction of all subsystems • Highest pump outlet pressures • Complex engine start-up due to multiple combustion devices	• Cycle limits chamber pressures to approx. 100 bar • Complex development effort due to close interaction of all subsystems • High pump outlet pressures

namely the RD-270 engine. A more detailed description of typical engine cycles is given in the following sections.

Booster Engines

Booster engines are ignited at sea-level conditions and operate in a duration of 1–3 minutes with usually high thrust requirements (up to 8000 kN). The specific impulse values achieved are in the low range, between 120 (shuttle) and 295 seconds for solid systems and up to 330 seconds for liquid systems. Some pertinent data can be found in Table 3.3.3.

Main Stage Engines

Main stage rocket engines operate at varying ambient conditions, being ignited on the launch pad at sea-level conditions and shut off in vacuum. Compared to booster engines, main stage engines operate for between 400 and 500 seconds with a lower thrust level (~ 2000 kN) but significantly higher specific impulse values (> 420 seconds for LOX/hydrogen). Table 3.3.4 contains characteristic values for realized and flight operational rocket engines. Figure 3.3.1 illustrates the Vulcain 2 engine as a typical first stage propulsion system, namely the core stage engine of the Ariane 5 rocket. The nozzle extension includes the turbine exhaust gas injector, designed to realize two objectives: film cooling of the divergent nozzle wall downstream from the injection location using the relatively cooler turbine exhaust gases, and increase of specific impulse by improved turbine exhaust gas expansion (Figure 3.3.6). The combustion chamber itself is hidden by the piping, and the fuel and oxidizer turbopumps are mounted on both sides of the combustion chamber.

Table 3.3.3: *Characteristic data for typical liquid booster engines.*

Type	Propellant combination	Thrust [MN]	I_{sp} [s]	Combustion chamber pressure [MPa]	Engine cycle
RD-170	LO_2/kerosene	7.65	310	25.1	Closed
RD-180	LO_2/kerosene	3.82	311	25.5	Closed
RD-107	LO_2/kerosene	0.81	257	5.9	Closed
F-1	LO_2/RP1	6.91	264	6.6	Open
MA-5A	LO_2/RP1	1.84	263	4.4	Open
RS-27	LO_2/RP1	0.91	263	4.8	Open
RD-253	N_2O_4/UDMH	1.47	267	14.7	Closed
YF-20	N_2O_4/UDMH	0.76	259	7.4	Open
Viking 6	N_2O_4/UH25	0.68	249	5.9	Open
RS-68	LO_2/LH_2	2.89	360	9.7	Open

Table 3.3.4: *Characteristic data for main stage liquid rocket engines.*

Type	Propellant combination	Thrust [MN]	I_{sp} [s]	Combustion chamber pressure [MPa]	Engine cycle
RD-108	LO_2/kerosene	0.78	248	5.1	Closed
RD-191	LO_2/kerosene	2.05	337	25.6	Closed
Viking C	N_2O_4/UH25	0.68	249	5.9	Open
SSME	LO_2/LH_2	1.67	364	20.5	Closed
RD-0120	LO_2/LH_2	1.45	354	20.6	Closed
LE-7	LO_2/LH_2	0.84	443*	14.5	Closed
Vulcain 2	LO_2/LH_2	1.0	433*	11.5	Open

* Vacuum impulse.

Upper Stage Rocket Engines

Upper stage rocket engines are ignited in vacuum (space) environment. These engines are characterized by much lower thrust levels (typically < 200 kN), but with requirements for very high specific impulse (> 460 s for H_2/O_2).

Table 3.3.5 includes characteristic data for flight-proven engines. Figure 3.3.6 shows the HM7B upper stage rocket engine used for the Ariane 5 ECA upper stage.

Apogee and Satellite Attitude Control Thrusters

Apogee and attitude control thrusters have stringent performance and reliability requirements to guarantee

Table 3.3.5: *Characteristic data for upper stage liquid rocket engines.*

Type	Propellant combination	Thrust [kN]	*Isp* [s]	Combustion chamber pressure [MPa]	Engine cycle
11D58M	LO_2/kerosene	84	355	4.4	Closed
RD-0210	N_2O_4/UDMH	582	327	14.8	Closed
Aestus	N_2O_4/MMH	30	325	1.0	PE
YF-75	LO_2/LH_2	79	440	3.7	Open
LE-5	LO_2/LH_2	122	452	3.8	Exp
HM7B	LO_2/LH_2	63	444	3.5	Open
Vinci	LO_2/LH_2	180	465	6.1	Exp
RL-10B	LO_2/LH_2	110	466	4.4	Exp

PE, pressure-fed engine; Exp, expander cycle (closed cycle) engine.



Figure 3.3.6: *Upper stage engine HM7B (Source: ASTRIUM).*

a service lifetime of up to 15 years in orbit, being the typical satellite life expectation. **Apogee thrusters** are designed for typical thrust levels of 500 N. **Attitude control thrusters** are designed for much lower thrust levels, around 10 N, with the unique requirement that they operate in both steady-state and pulse modes. Both thruster types are described in more detail in Section 4.4.

3.3.2.3 Solid Propulsion Engines

Solid propulsion engines are often selected when high reliability, quick operational readiness and simple storage are required. They are used as **gas generators** to start the turbopumps of liquid engines, as **acceleration rockets** for stage separation, propellant settling of upper stages, or as main or strap-on engines of the launch vehicle.

The main advantage of solid propulsion lies in its simple layout based on only a few components: motor case, internal thermal protection, solid propellant, nozzle and ignition system. The main disadvantage is the low specific impulse. Typical values of 250–295 s are about 40% below cryogenic systems. The dry mass is also relatively high, since the motor case serves as propellant tank and combustion chamber concŭrrently. Therefore, the case is designed for high internal pressure levels requiring corresponding wall thicknesses. The main use of solid propulsion systems is therefore the **high thrust domain**; for example, for first stages, the high–energy density of solid propellants allows high mass flow and therefore high thrust combined with relatively small volume. This is often more important than the specific impulse for first stages.

Propellant Components and Production

Solid propulsion motors are **single-component systems**; propellant and an oxygen containing chemical are mixed together to form a rubber-like block, which burns autonomously after ignition. Ariane 5 (see Figure 3.3.7) uses ammonium perchlorate (NH_4ClO_4) as oxidizer and hydroxyl-terminated polybutadiene (HTPB) rubber as propellant, with aluminum powder to increase gas temperatures and the specific impulse. A curing agent is added to promote solidification of the rubber/oxidizer mixture.

Propellant Production

The different components of the propellant are weighed and mixed in large batches (about 10 t of

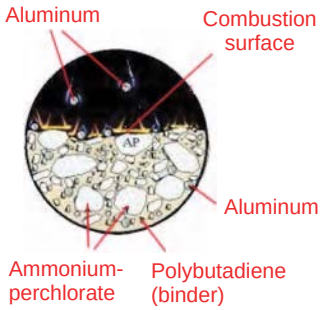


Figure 3.3.7: *Propellant components.*

propellant per mixer). The mixing process produces a homogeneous semiliquid propellant with honey-like viscosity. It is then transferred to a casting pit where it is poured into the motor case, which is equipped with a casting mandrel precisely determining the combustion channel. The casting process is performed under vacuum conditions in order to avoid the formation of bubbles in the propellant block. After complete loading of the motor case, the polymerization process of the propellant is initiated by increasing the temperature within the casting pit. This curing cycle allows the propellant to solidify. After the end of the curing process, as the temperature in the pit decreases the propellant shrinks, allowing the mandrel to be extracted.

Burning Types

The thrust curve is determined by how the combustion surface develops over time. Burning types are divided into the following categories (see Figure 3.3.8).

Neutral burners are frontal burners (like cigarettes) with a constant circular combustion surface. They generate constant thrust levels by constant mass flows. The hot combustion gases are in contact with the motor case from the beginning of the burn. Consequently, long burn durations would require excessive internal thermal protection layers and increase weight to unacceptable levels. Therefore, frontal burners have limited burn durations and are mostly used for stage separation and propellant settling tasks.

Progressive burners combine frontal burners with cylindrical inner channels. The combustion surface of the inner channel widens during the burn, which

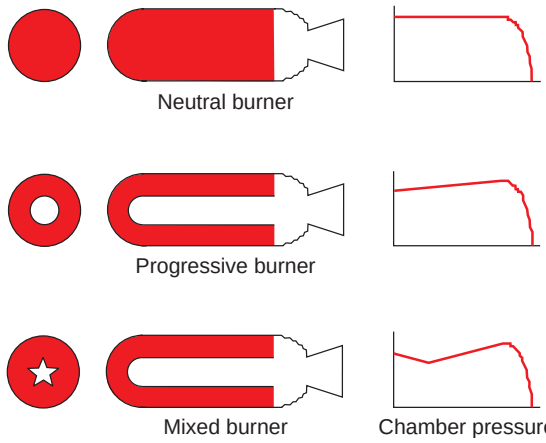


Figure 3.3.8: *Burn types of solid propulsion motors*.

steadily increases the generated thrust. This burner type is used mostly for upper stage applications or solid apogee motors.

Mixed burners allow an optimum adaptation of the thrust curve to the needs and constraints of the selected trajectory. A combination of star-shaped burners with cylindrical or conical blocks can provide an excellent basis for first stages: the important combustion surface of the star provides high thrust for liftoff; after burnout of the star surface the combustion surface and the resulting thrust decrease. Dynamic pressure levels, accelerations and gravitation losses remain in the predefined limits. Figure 3.3.9 provides a view of the frontal segment of the Ariane 5 booster with its typical star shape.

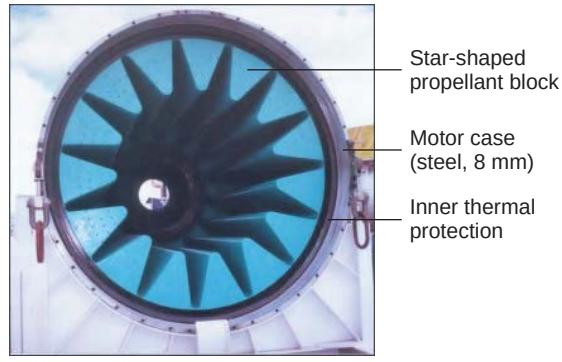


Figure 3.3.9: *Forward segment of Ariane 5 booster (Source: ESA).*

Equation 3.3.6 describes the important link between **mass flow** and resulting thrust and can be adapted for solid propulsion motors as follows:

$$\dot{m} = \rho A_b v_{ab} \qquad (3.3.6)$$

where:

ρ = propellant density,
A_b = combustion front surface,
v_{ab} = combustion front velocity.

An analysis of equation (3.3.1) and equation (3.3.6) shows that the generation of a high liftoff thrust requires high propellant densities, high combustion surfaces, high burning velocities and high combustion pressures. Each of these parameters has its own limits.

High pressure levels lead to high dry masses; large combustion surfaces reduce the propellant filling level of the motor (= internal volume/propellant volume) and increase the dry mass. The **combustion velocity** is also limited: velocities that are too high increase the risk of erosive burning. In this case the combustion gas flow passage grows to a conical shape toward the end of the propellant segments thereby exposing the lower booster case structure to the full heat load early on.

Ariane 5 Solid Booster

The two Ariane 5 solid boosters provide approximately 13 MN, more than 90% of the liftoff thrust. Each booster is built from three segments. The dry mass of one booster is about 31 t and the propellant mass is about 240 t. The burn time is 132 s and the boosters accelerate the launcher to 2 km/s at 69 km altitude. Figure 3.3.10 shows a booster in Kourou, on the left as a drawing and on the right after rollout from the booster assembly building.

The **motor case** consists of 8 mm thick steel and is insulated from the hot combustion gases by an internal thermal protective layer up to 12 mm thick.

The **nozzle**, see Figure 3.3.11, is built of composite materials and has a gimbal joint allowing orientation of the thrust vector for launcher control. The following environment had to be taken into account for the nozzle design: temperatures in the region of the nozzle throat reach 3000 °C; the combustion gases

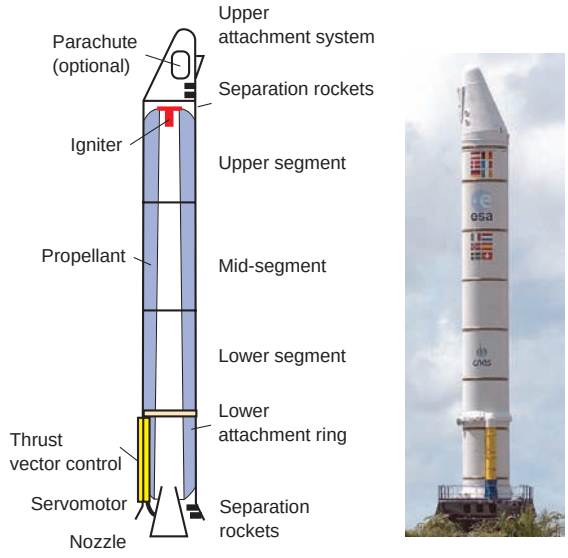


Figure 3.3.10: *Ariane 5 booster (Source: ESA).*

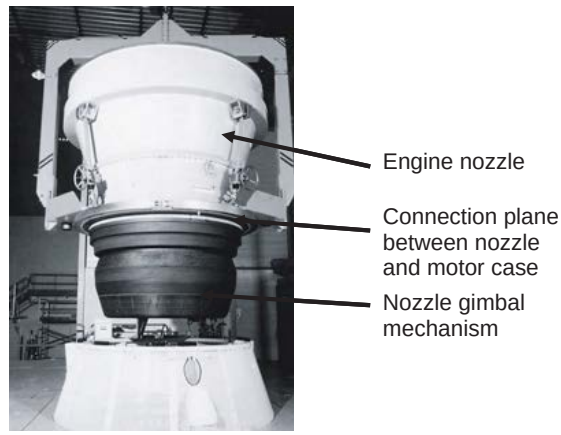


Figure 3.3.11: *Engine nozzle (Source: ESA).*

pass the nozzle at a rate of 2 t per second and contain aluminum or aluminum oxide particles which can produce an abrasive effect in case of contact with the nozzle's internal surface, given the high velocities of the particles; the inner diameter of the nozzle throat widens during the burn time by several millimeters through erosion; and the nozzle must allow swiveling of 7° to allow for thrust vector control.

The swiveling moment should not be too high in order to limit the energy needed. The swiveling mechanism has to be gas tight to prohibit leaks of

combustion gases. This leads to a rather complex composite gimbal joint design employing accordion-like outer walls.

3.3.3 Engine Components

The main components of a rocket engine, the gas generator, the turbopump and the thrust chamber assembly with propellant manifolds, injection head, ignition system, combustion chamber and nozzle, are strongly inter-connected functionally and physically.

3.3.3.1 Injector Head

The main objectives of the injection head are to supply and homogeneously distribute the propellants and to decouple the engine and thrust chamber subsystems in order to avoid both mechanical vibrations and combustion instability. Besides damping and homogenizing the secondary flows which result from the turbomachinery, the pressure losses of the injection elements decouple almost entirely the propellant manifolds and the combustion chamber. Typical losses amount to about 15–25% of the combustion chamber pressure. In addition to the injector elements and the propellant manifolds, the injector head typically also includes an appropriate ignition system.

Engine Start-up and Ignition System

The engine start-up and ignition system is one of the most critical systems of a rocket engine. Malfunctions during start-up and ignition are among the most likely reasons for launch failures. Hence, the thrust chamber, ignition system and start-up sequence have to be designed and developed in parallel to ensure safe and stable operation.

The ignition system must provide the necessary energy at the right location and time, long enough to ignite the propellants injected; an example of a typical pyrotechnic igniter is shown in Figure 3.3.12.

In order to ensure this function it has to fulfill a number of requirements. First, the mixture ratio of the propellants has to be near stoichiometric and the overall conditions have to be favorable for flame spreading at the injectors. Second, the energy

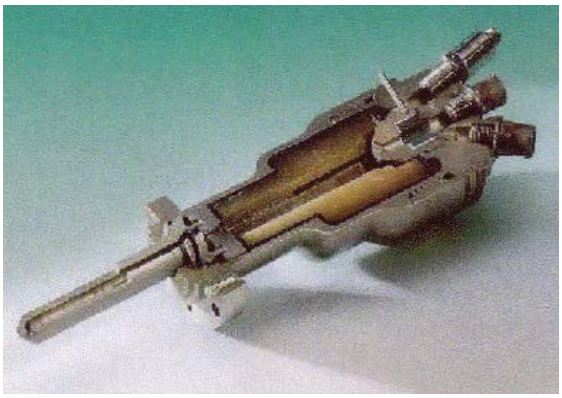


Figure 3.3.12: *Pyrotechnic igniter of the Vulcain 1 engine (Source: ESA).*

provided has to be sufficiently high to ensure propellant vaporization and heating propellants to temperatures which exceed the ignition temperature. Propellant temperatures under 100 K make it obvious why heat transfer is a key element of the engine start-up. An ignition delay of only a few milliseconds may be sufficient to accumulate enough propellants which, when they react, yield pressure peaks which can be harmful to the combustion chamber or for other components of the engine. Additionally, such pressure peaks can trigger combustion instabilities which almost always lead to damage of the engine and a loss of the entire mission. The arguments mentioned above make it clear that an ignition system strongly depends on the engine start-up and thus has to be developed in parallel with the engine and matched closely with the injection head and the thrust chamber.

Injection Elements

There are four different principal types of injector head elements, depending on the propellant combination and the application: swirl, impinging, showerhead and coaxial injection.

Swirl Injection

In most cases, the injection element openings for the oxidizer and the fuel are arranged coaxially to produce significant tangential injection velocity components for both propellants. Figure 3.3.13 depicts a typical example of such an injection element.

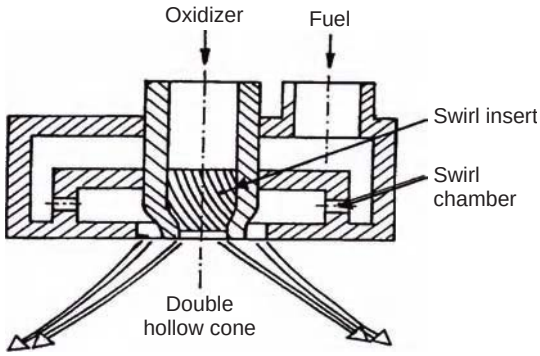


Figure 3.3.13: *Swirl injection.*

Swirl injection results in good propellant mixing and combustion in combination with an induced recirculation zone in the center of the double hollow cone. An additional advantage of the swirl injection element type utilized in small thrusters is the generation of a cooling film along the chamber wall even without a dedicated film cooling design feature. Drawbacks of swirl injection elements are the sensitivity to manufacturing intolerances and the strong interference with the combustion efficiency and the thermal wall loading.

EADS Astrium employs this kind of injection element technology in all small bipropellant thrusters used for satellite positioning and attitude control in the thrust range from 10 to 500 N. In Russia, swirl elements are also utilized for multielement injection systems in high-thrust liquid rocket engines.

Impinging

Impinging injection is mainly used for propellant combinations with low vaporization rates, for example MMH/N_2O_4 in combination with short combustion chamber lengths. Propellant mixing is accomplished by impinging the different propellant components at an angle of about 120°. The number and geometrical positioning of the injection holes determine the type of impinging, such as unlike impinging (two different propellants mix) and like impinging (the same propellant mixes), as well as doublets, triplets, quadlets and pentads. Figure 3.3.14 depicts the different impinging types.

As with swirl injection, impinging injection technology is also sensitive to manufacturing intolerances and carries the risk of large and heterogeneous heat loads on the combustion chamber wall as well as on the injector faceplate. In addition, impinging injection technology is also very sensitive in that it can lead to combustion instability.

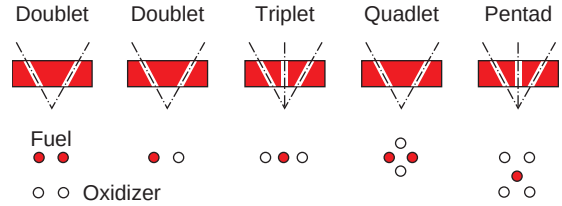


Figure 3.3.14: *Various impinging injection types.*

Applications of the impinging element technology can be found in the rocket engines of the Thor, Atlas, Saturn (for the propellant combination LOX/kerosene) and Titan II launchers, and the Moon landing engine which utilizes N_2O_4/UDMH-N_2H_4.

Showerhead

Showerhead injection is associated with the least manufacturing effort and features the lowest thermal load conditions of the injector head faceplate and combustion chamber wall. The drawback of the showerhead is a nonoptimized mixing process of the propellants and the resulting lower combustion efficiency, since the propellants are injected only axially and the mixing itself depends on the turbulence occuring in the combustion chamber.

In principle, showerhead injection element technology consisting of two or several injection holes can be applied to all liquid propellant combinations and various showerhead configurations.

Coaxial Injection

The mixing of the propellant components using coaxial injection technology is by direct contact of two tubular jets of propellant. The high-velocity differential of the two jets results in large shear forces imposed on the two propellants. Besides good mixing behavior, coaxial injection element technology is characterized by low functional interference of the injection elements, good scalability in terms of number of elements of the injector head, and low thermal load conditions of the injector head faceplate. Figure 3.3.15 depicts

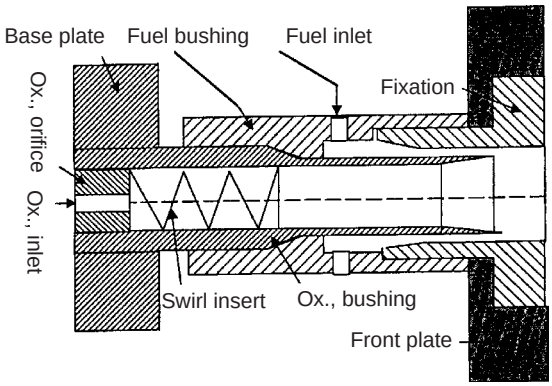


Figure 3.3.15: *Coaxial injection.*

Table 3.3.6: *Typical coaxial element operating conditions: typical data for coaxial injection for cryogenic and storable propellant combinations.*

Cryogenic Propellent Combination		Storable Propellent Combination	
H_2 velocity	300 m/s	MMH velocity	14 m/s
LOX velocity	15 m/s	NTO velocity	12 m/s
Element loading	500 g/s	Element loading	70 g/s
Diffuser exit angle	6°	Diffuser exit angle	0°
Recess	3.5 mm	Recess	3.5 mm

this technology and Table 3.3.6 lists typical operating conditions for coaxial injection elements.

Key advantages of coaxial injection element technology are very good scalability and excellent combustion efficiency. In addition, manufacturing reproducibility is high and therefore the scrap rate is low. These key advantages make coaxial injection element technology one of the preferred injection methods for LOX/LH_2 and LOX/methane liquid rocket engines.

Figure 3.3.16 depicts the injector head of the Vulcain liquid rocket engine including igniter, LOX distributor, the base plate with the injection elements, the faceplate and the acoustic absorber featuring Helmholtz resonators. Additional examples are the Space Shuttle main engine, the Delta 4 liquid rocket engine RS-68, and the cryogenic upper stage expander cycle engine Vinci, currently under development.

An appropriate combination of swirl elements for the oxidizer with impinging elements for the fuel allows application of coaxial injection element technology not only for cryogenic propellant combinations, but also for storable propellant combinations, as was designed for the upper stage Aestus engine in the Ariane 5G launcher.

The achieved combustion efficiency depends mainly on the propellant preparation and mixing, and in particular on the atomization. The dominating physical phenomena are determined by the velocity ratio of the injected propellants, the inertia–viscous

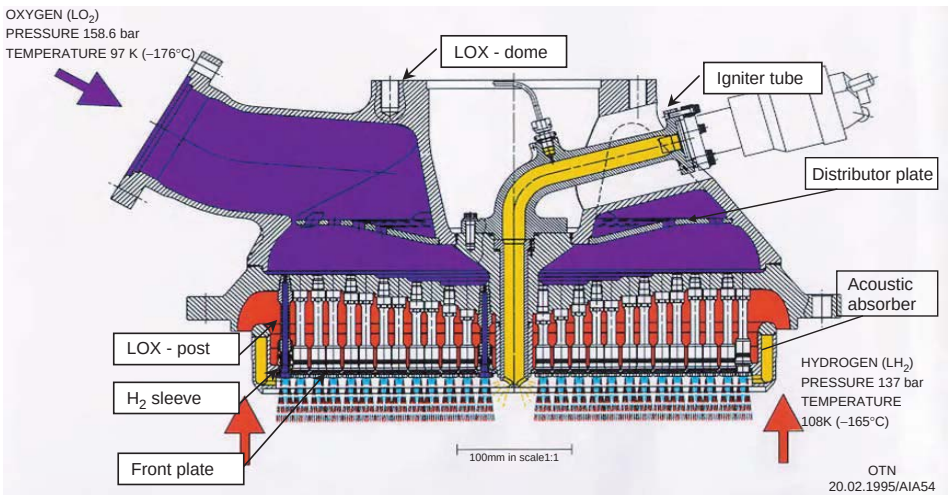


Figure 3.3.16: *Injector head of the Vulcain engine.*

force ratio (Reynolds number), the inertia–surface tension ratio (Weber number) and the impulse ratio of the injected propellants (J number):

$$\frac{u_g}{u_1}; \quad Re = \frac{\rho d u}{\mu}; \quad We = \rho_g \frac{(u_g - u_1)^2}{\sigma}; \quad J = \frac{(\rho u^2)_g}{(\rho u^2)_1} \tag{3.3.7}$$

An important effect influencing the atomization process is that the oxidizer is injected below its critical temperature but above the critical pressure. Thus, thermodynamic effects influence the atomization near the injector head.

Development Logic

The development logic is driven by the following verification methods: analysis, including similarity and comparability; testing; inspection; and review of design. Verification by testing is certainly the most expensive and was in past years the most preferred method to demonstrate the reliability of a rocket engine system. Recent improvements in the numerical simulation of complex physical phenomena present in rocket engines are shifting the verification method from test to analysis. Today's modeling capability includes the representation of complex combustion and flow phenomena, which allows a high level of predictability of engine behavior. In addition, a large amount of testing adversely affects the development costs, which is nowadays no longer justifiable. Future rocket engine developments will further change the development logic such that a larger extent of the verification process will be increasingly performed by means of analysis.

Rocket engine development programs will always depend on hot fire tests, although the analytical verification method will become stronger, as mentioned above. A balance between demonstrated reliability level and available development budget must be found. An approach relevant for the development of an injector head is depicted in Figure 3.3.17.

Hot fire tests are performed on the element level using relevant scale injector head and combustion chamber models, keeping the size of the injection element unchanged. These test specimens are equipped with a sufficiently large number of thermoelements, static and dynamic pressure sensors, acceleration meters and strain gauges in order to calibrate the simulation models and consequently improve the level of predictability. EADS Astrium has used this approach for several years with great success, as demonstrated on a cryogenic upper stage expander thrust chamber. The temperature increase inside the cooling channels of the combustion chamber was within 2% at an absolute temperature level of 250 K.

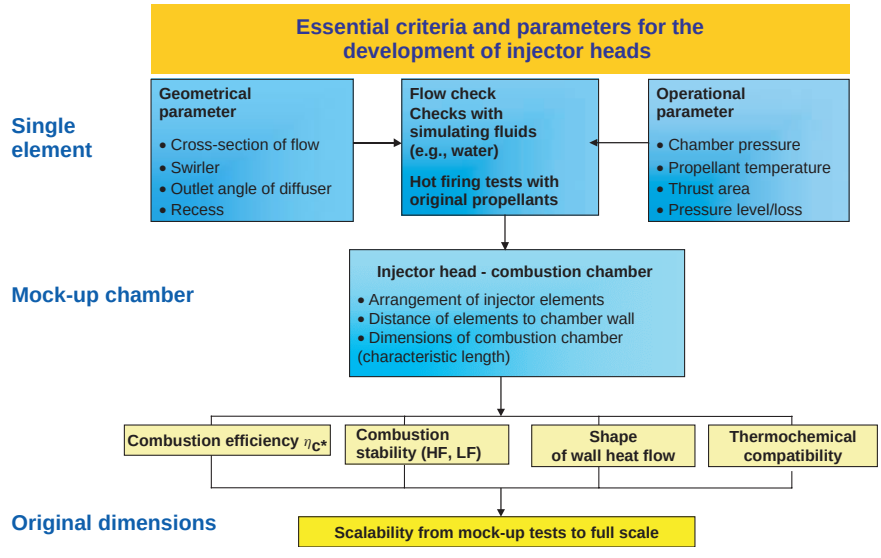


Figure 3.3.17: *Injector head development methodology.*

Other performance characteristics are of the same order of magnitude.

In addition to the efforts to improve analysis methods, consistent improvement in the design of experimental methods for hot fire tests is under consideration, not only to reduce the number of tests, but also to increase the level and amount of information gained, in order to further optimize the development logic used at EADS Astrium.

3.3.3.2 Combustion Chamber and Nozzle

Besides the turbomachinery and the injection head, the most important components of a rocket engine are the combustion chamber and the nozzle. The combustion chamber consists of a relatively short cylindrical part, followed by a throat area with a gradually converging part, the throat section having a contraction ratio of about 2.5 where the highest thermal loads in the chamber occur. The diverging part of the chamber typically extends down to expansion ratios of 5–8 and is integrated into the combustion chamber using similar materials and cooling philosophy. The thrust nozzle itself is a separate component often fabricated using a different material as well as a different cooling cycle, namely film cooling, dump cooling or radiation cooling. The thrust chamber assembly (TCA) subsystem consists of injection head, ignition system, combustion chamber and nozzle.

Combustion Chamber

The main objective of the combustion chamber (Figure 3.3.18) is to completely burn the propellants and to accelerate the exhaust gases to sonic velocities in the throat. Design difficulties relate to precise and reliable predictions of an optimum liner contour, combustion efficiency, hot gas side and coolant side heat transfer and appropriate cooling, required component lifetime and, finally, reliable and justifiable requirements for and interface conditions between the TCA components.

Cooling of Combustion Chamber and Nozzle

The key challenge of the entire combustion chamber and nozzle design process is to predict cooling system behavior and performance. A collection of different cooling processes and possibilities is shown



Figure 3.3.18: *Combustion chamber of Vulcain 2 (Source: ASTRIUM).*

in Figure 3.3.19, using for illustration a combustion chamber wall cross-section including a cooling channel and pertinent heat transfer mechanisms.

The standard cooling method in rocket engines is regenerative cooling. One propellant, typically the fuel, enters through a distribution manifold at the combustion chamber end and is directed in a counterflow direction upstream through cooling channels to the injector head and thus enters the combustion chamber through appropriately designed injectors. Both the hot gas side as well as the coolant side heat transfer, which define the overall thermal loads and the cooling channel pressure loss, are extremely dependent on the cooling channel design. Examples

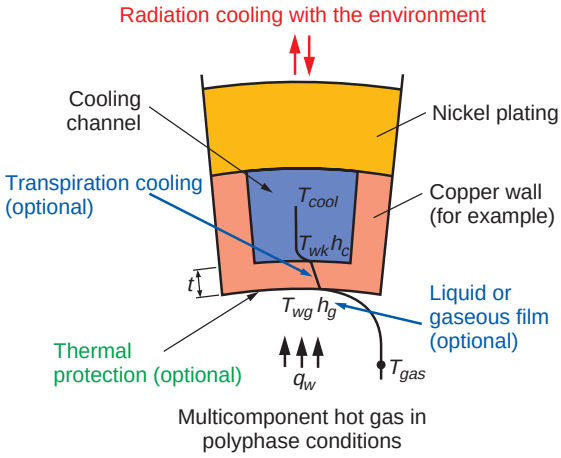


Figure 3.3.19: *Heat transfer mechanisms.*

of typical cooling channel designs are presented in Figures 3.3.20 and 3.3.21. While Figure 3.3.20 shows the raw copper liner with milled cooling channels, Figure 3.3.21 shows a cross-section of a Vulcain-type liner for the combustion chamber wall with the copper cooling channels and outer shell made of electrically deposited nickel.

Hot gas side and coolant side heat transfers are coupled via the heat conductivity in the liner structure and hence the entire heat transfer problem can only be solved in a fully coupled manner. However, such a coupled solution based on numerical results is not to be expected in the near future. The reasons are both numerous and serious: the differences in length scales of the combustion chamber, injector element and boundary layer; the time scales of nonequilibrium thermodynamics; finite rate chemistry; the presence

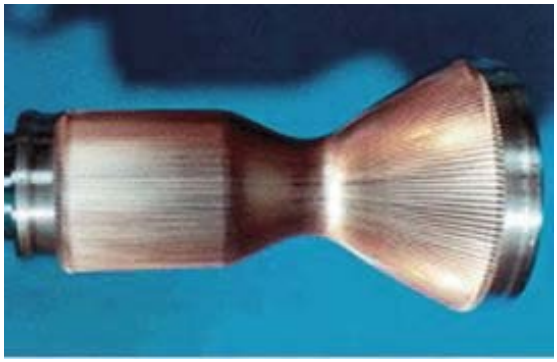


Figure 3.3.20: *Copper liner with milled cooling channels (Source: ASTRIUM).*



Figure 3.3.21: *Combustion chamber liner cross section (Source: ASTRIUM).*

of areas with subsonic, transonic and supersonic velocities; atomization in general and atomization under subcritical, transcritical and supercritical conditions in particular; and last but not least, the necessity for complex thermodynamic descriptions of processes, for example the properties and behavior of gases, liquids and solids under cryogenic conditions. Furthermore, dissociation of the exhaust gases has a direct impact on the combustion efficiency but may also cause a local temperature increase in the cooler boundary layer due to recombination. Finally, catalytic reactions at the surface may additionally influence the local heat balance and thus have an impact on the overall heat transfer. In the case of hydrocarbon fuels, decomposition reactions due to pyrolysis in the cooling channels may further increase the complexity of the coupled problem.

With all that said, it is rather obvious that semi- or fully empirical correlations in the form of Nusselt relations are used; a typical example is the Bartz equation, which describes the hot gas side heat transfer:

$$Nu = 0.062\, Re^{0.8}\, Pr^{0.3} \qquad (3.3.8)$$

$$(10^7 < Re < 10^8,\ Pr \approx 0.5)$$

Various modifications of this basic relation exist, generated by engineers trying to make use of known or measured local geometrical quantities or thermodynamic and fluid properties which have an influence on the local heat transfer in order either to improve the predictive capabilities of the relation for a given set of operating conditions, or to enlarge the parameter range of application:

$$\alpha_g = \left[\frac{0.026}{D_t^{0.2}}\left(\frac{\mu^{0.2}\, c_p}{Pr^{0.6}}\right)_{ns}\left(\frac{(p_c)_{ns}\, g}{c^*}\right)\left(\frac{D_t}{R}\right)\right] \cdot \left(\frac{A_t}{A}\right)^{0.9} \sigma \qquad (3.3.9)$$

with

$$\sigma = \left[\frac{1}{2}\frac{T_{wg}}{(T_c)_{ns}}\left(1 + \frac{k-1}{2} M\, a^2\right) + \frac{1}{2}\right]^{-0.68} \cdot \left[1 + \frac{k-1}{2} M\, a^2\right]^{-0.12} \qquad (3.3.10)$$

$$Pr = \frac{4\,k}{9\,k-5} \quad \text{and} \quad \mu = (46.6 \cdot 10^{-10})\, M^{0.5}\, T^{0.6} \qquad (3.3.11)$$

Simple correlations (Equation 3.3.12) are in use to predict the coolant side heat transfer:

$$Nu = K\, Re^a\, Pr^b \qquad (3.3.12)$$

The dimensionless parameters such as the Reynolds or the Prandtl numbers, see Table 3.3.7, and their exponents vary according to the coolant used.

A table with coefficients reflecting the influence of combustion chamber pressure, different temperatures, heat flux densities and fluids which are used for much more sophisticated correlations of the form

$$Nu = K\, Re^a\, Pr^b \left(\frac{T_w}{T_b}\right)^n \left(1 + \frac{2\,D}{L}\right)^m \qquad (3.3.13)$$

or as

$$Nu = K\, Re^a\, Pr^b \left(\frac{\rho}{\rho_w}\right)^c \left(\frac{\mu}{\mu_w}\right)^d \left(\frac{k}{\kappa_w}\right)^e \cdot \left(\frac{\bar{c}_p}{c_p}\right)^f \left(\frac{p}{p_{cr}}\right)^g \left(1 + \frac{2\,D}{L}\right)^m \qquad (3.3.14)$$

Table 3.3.7: *Coefficients of the Bartz relation for coolant side heat transfer.*

	K	a	b	
CH_4	0.0023	0.8	0.4	Re ~ 10^6–10^7, Pr ~ 1
Kerosene	0.005	0.95	0.4	$Re < 2 \cdot 10^4$
	0.023	0.8	0.4	$Re > 2 \cdot 10^4$
LH_2	0.62	0.8	0.3	$Re \approx 10^6 - 10^7$, $Pr \approx 1$

is provided in the appropriate specialized literature quoted at the end of this section. All the coefficients and exponents of Equations 3.3.9 to 3.3.14 try to quantify the impact of

- Geometry (cooling channel dimensions and their change along the combustion chamber length, curvature)
- Chemistry (finite rate chemistry, catalytic wall effects, dissociation and recombination in the boundary layer, pyrolysis of the coolant)
- Thermodynamics (real gas behavior, near-critical behavior, varying fluid properties)
- Fluid dynamics (turbulence, atomization, mixing, stratification)

on either the hot gas or coolant side heat transfers. All these correlations are based on different experiments from different working groups and, therefore, the results and finally the coefficients obtained depend on the experimental setup, the facilities and operating conditions and the measuring techniques applied, and accordingly include all the known and unknown errors as well.

Film Cooling

Film cooling as the sole cooling method is applied only in satellite engines due to its limited efficiency. A special type of film cooling is "injector trimming," where the outer elements are operated with a mixture ratio that achieves an appropriate combustion chamber wall cooling. Regenerative cooling applying tubes or milled cooling channels reaches its limits at heat flux densities of about 30 MW/m^2 and 80 MW/m^2, respectively. At higher heat flux levels it is usually assisted by film cooling, examples being the Vulcain 2, SSME or the RD-180 engines. The two general key problems of film cooling design are film stability and cooling efficiency. Equation 3.3.15 may be seen as a general relation for film cooling design:

$$\frac{T_{aw} - T_{wg}}{T_{aw} - T_{co}} = e^{-\left(\frac{\alpha_g}{F\, c_{pc}\eta_c}\right)} \qquad (3.3.15)$$

Aside from the cooling effect of films entering at the injector head, they also protect the wall from aggressive gases. This cooling method is, for example, applied for SSME and Vulcain 2 engines. The

RD-171 engine applies two sorts of films, one directly at the injector head and another one in front of the converging part of the throat to reduce thermal loads in an area where peak heat fluxes are to be expected. All the chambers of these engines are in addition regeneratively cooled.

Transpiration Cooling

Transpiration or effusion cooling can also be seen as a special type of film cooling where an appropriate part of the fuel is injected into the combustor through a porous wall in order to establish the necessary cooling. While sintered metals offer only limited advantages since the maximum wall temperature stays the same and the wall thickness has to be increased to maintain the mechanical integrity of the structure, fiber-reinforced ceramic matrix composites offer the possibility to reduce the structural weight and the necessary coolant mass as well, since they allow for much higher wall temperatures:

$$\frac{T_{aw} - T_{co}}{T_{wg} - T_{co}} = [1 + \{1.18\ (Re_b)^{0.1} - 1\} - Y]\, Y^{Pr_m} \qquad (3.3.16)$$

with $Y = e^{37\left(\frac{G_c}{G_g}\right)(Re)^{0.1}}$

Ablative Cooling

This cooling method may be seen as a special form of film cooling and has been applied either in the throat of medium-pressure, short-burn-time booster engines or in low-pressure satellite engines. The functional principle is based on the interaction of the following phenomena: heat input from the hot gas flow leads to melting and vaporization of the wall material and establishes together with an endothermic reaction a near-wall coolant film. Materials used are carbon, C/C or SiC structures without infiltrated hydrocarbons. Those hydrocarbons may also have favorable additives incorporated, such as tungsten (W) or rhenium (Re). The disadvantage of the method lies in its limitation of the burn time and operating regime of the engine. Quite often ablative cooled systems additionally apply film cooling in the injection area in order to reduce injector/wall interaction and reduce the heat flux in general. As an example, both the Viking and the RS-68 engines have ablative cooled nozzle throats. Common characteristics of both engines are short burn times and comparatively low chamber pressures of 205 and 250 s, and 59 and 97 bar, respectively. Any design is entirely dependent on the materials and additives applied and the operating conditions of the combustor. A specific mathematical relation is omitted here. The only thing to be mentioned is that the applied design logic can be based on that of a solid booster; however, detailed parameters representing the specific characteristics of the materials used have to be determined experimentally.

Radiation Cooling

Aside from the conventional closed cooling cycles applied in combustion chambers, thrust nozzles are cooled either in an open cycle mode called "dump cooling" or by radiation cooling. Obviously, this method relies on materials withstanding high temperatures and thus is only applicable for satellite engines; however, it is often used for the extensions of thrust nozzles. Independent of the material applied, be it refractory metals such as tungsten, rhenium or iridium, or ceramic matrix composites, the surface of thrusters working within the atmosphere has to be covered with an oxygen protective layer. Upper stage engines such as the RL10-B or the new Vinci usually use a ceramic matrix composite (CMC) nozzle extension. The design of a radiation-cooled device follows Equation 3.3.17 with the emission coefficient ε, the Stefan–Boltzmann radiation constant σ and the maximum wall temperature T_{wg}:

$$q = \varepsilon\, \sigma\, T_{wg}^4 \qquad (3.3.17)$$

It is worth mentioning that radiation-cooled thrusters or nozzles require a shielding system which protects sensitive engine parts and measurement equipment from high heat loads.

Nozzle Extension

The main function of the nozzle extension is to optimally expand the hot combustion gases and to maximize the exhaust velocity to create maximum thrust (see Equation 3.3.1). Bell-shaped nozzles with a parabolic contour (see Figure 3.3.22) generate an exhaust flow of high axial parallelism at minimal losses.

In general two kinds of contours can be distinguished:

1. Classical ideal nozzles, with contours similar to wind tunnel nozzles and lengths adapted to the ambient pressure.
2. Thrust-optimized nozzles, their contours opening at somewhat steeper angles behind the throat, thus leading to shorter length and less mass at the same expansion ratio.

Thrust-optimized contours are more sensitive to side loads during operational transients because of their internal specific fluid conditions and therefore necessitate more robust structures compared to ideal nozzles.

One of the basic problems of booster engine nozzles results from their long burning time. The exit pressure of the nozzle on the ground (during start) is smaller than the ambient pressure (overexpansion), and therefore the thrust is lower than optimal. If the expansion ratio exceeds a certain limit, flow separation occurs in the nozzle extension, creating destructive side loads. During the ascent the ambient pressure decreases steadily until the optimum thrust is reached when ambient and exit pressures are identical. During further ascent the ambient pressure continuously decreases below the exit pressure (underexpansion) and deviation from the optimum thrust increases. The losses at the start and end of the burn time can reach 10 to 15%, as Figure 3.3.23 shows.

The design of the nozzle is therefore always a compromise between safety during ground operation and achievable specific impulse at altitude. A remarkably higher specific impulse could be achieved if a continuously adaptive nozzle were technically feasible, as Figure 3.3.23 illustrates. However, three possibilities exist to at least partially realize the illustrated potential, which are:

1. The so-called dual bell, featuring transition from sea level to ascent altitude conditions at distinct ambient pressure, see Figure 3.3.24(a).

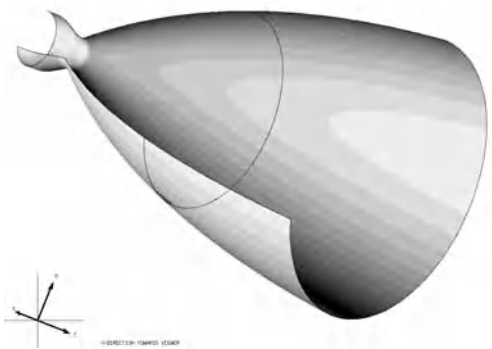


Figure 3.3.22: *Typical bell-shaped nozzle.*

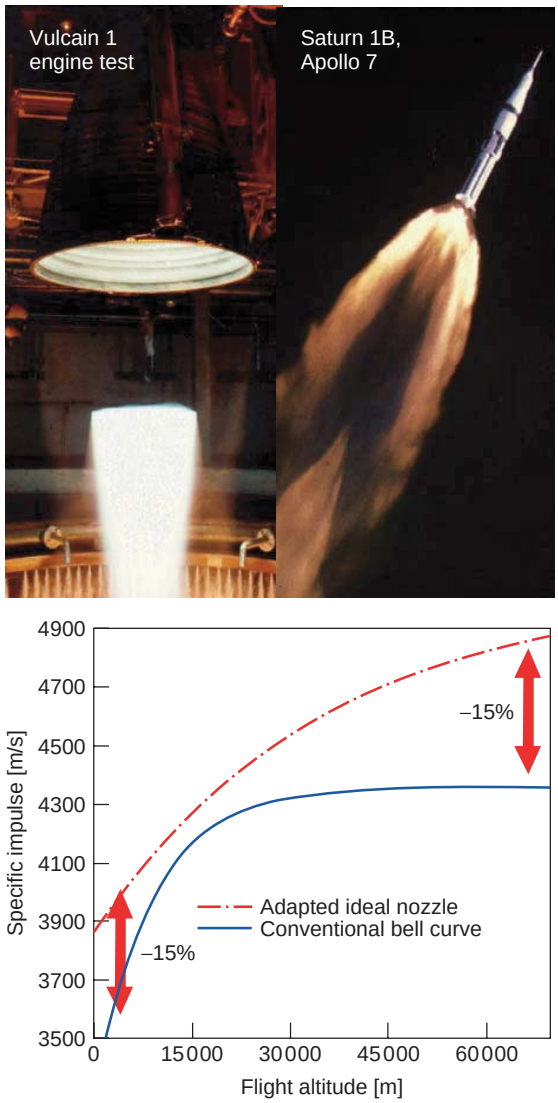
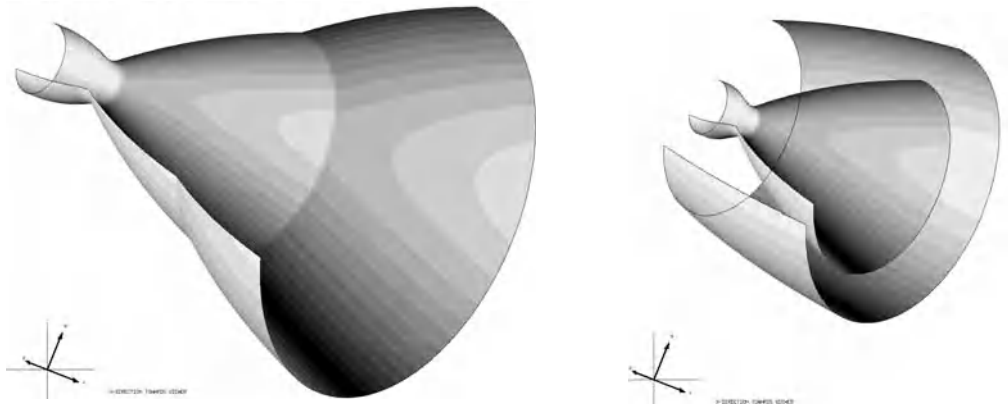


Figure 3.3.23: *Influence of a nonadapted nozzle on specific impulse as a function of altitude (Source: NASA/DLR).*



(a) (b)

Figure 3.3.24: *Dual bell (a) and extendable nozzle (b).*

2. An extendable nozzle, which can be elongated during operation at higher altitudes, see Figure 3.3.24(b).
3. An internal nozzle cone which is dropped at the appropriate altitude, see Figure 3.3.25.

While the dual-bell nozzle, only suitable for first-stage booster engines, has so far not been operated during flight, the upper stage RL-10B and Vinci engines feature extendable nozzles, but for the purpose of reducing the height and the weight of the launcher. These nozzles are extended after the first stage is dropped and prior to the ignition of the upper stage engine.

In principle two kinds of nozzle design are used in today's rocket engines: a variant built up by tubes which are joined by welding or brazing (see Figure 3.3.26), or an integral version in which the cooling channels are milled into an inner cone, covered by an outer cone joined to the inner one by welding or brazing, as shown in Figure 3.3.27.

Both variants are realized as closed as well as open engine loops. The first necessitates the redirection of the nozzle coolant into the combustion chamber,

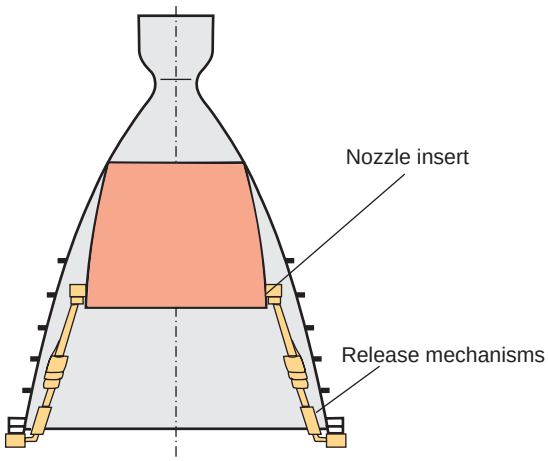
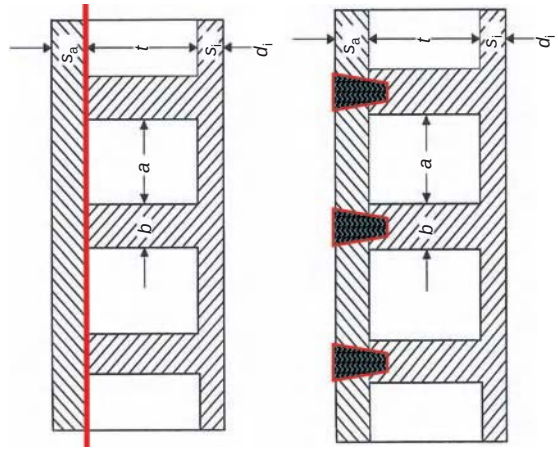


Figure 3.3.25: *Nozzle with discardable inner cone.*



(a) (b)

Figure 3.3.26: *Design principles of cooled nozzles.*



Figure 3.3.27: *Nozzle with cooling channels.*

while in the second the coolant is dumped at the nozzle exit.

Design of Rocket Combustion Chambers

The design of a rocket combustion chamber mostly depends on how the chamber is cooled. Engines with short burning times associated with low to only moderate heat fluxes allow simple steel design chambers which are either specially coated and film cooled, or contain ablative inserts which slowly burn away and in the process cool the chamber. The thrust chamber of the Viking engine is often quoted as a typical example of such a design. Higher chamber pressures, more energetic propellant combinations and longer burning times require the above-described regenerative cooling and accordingly a much more complex design. The design of such combustion chambers is usually characterized by two different functions realized in two main parts:

1. The inner part of the combustion chamber, the so-called combustion chamber liner, which is made of a material having high heat conductivity, such as a copper alloy.
2. The outer part of the combustion chamber, which provides the load-carrying structure.

These two parts are illustrated in Figure 3.3.21 above. The cooling channels are integrated in the liner material. A typical material combination is based on Narloy Z (copper–silver–zirconium alloy) for the liner and high-strength nickel for the load jacket.

The state-of-the-art **manufacturing technology** for high-performance engine combustion chambers is as follows.

The chamber liner is forged to a high deformation degree and has milled cooling channels. These are then filled with wax to prevent their blockage during the subsequent manufacturing process and protected by a covering layer of thin galvanic copper. The load-carrying nickel structure is subsequently deposited by a galvanic process. After finalization of the galvanic deposition the wax is expelled from the cooling channels by moderate heating. Inlet and outlet manifolds for the coolant are joined to the jacket by EB (Electron Beam) welding. This manufacturing technology was developed and patented by MBB, the predecessor of today's EADS Astrium Space Transportation, in the 1960s. The combustion chambers of the SSME and also of Vulcain 1 and Vulcain 2 are built using the method described above.

As an alternative the electro-deposited jacket structure can also be brazed to the combustion chamber liner. This technology is widely used, especially in Russian engines. But also the relatively new RS-68 engine features a combustion chamber built using brazing technology.

3.3.3.3 Gas Generators

The main purpose of a gas generator or, in the case of a staged combustion cycle, the engine preburner is to provide the necessary hot gases to drive the turbopump. The **power requirement of the turbine** depends on the mass flow rate and the thermodynamic properties (c_p, κ, T_1) of the driver gas, on the turbine efficiency and on the available pressure ratio p_2/p_1:

$$P = \eta\, \dot{m}\, c_p\, T_1 \left[1 - \left(\frac{p_2}{p_1} \right)^{\left(\frac{k-1}{k} \right)} \right] \qquad (3.3.18)$$

While the turbine exit pressure p_2 of bypass or gas generator cycle engines is not directly coupled to the pressure in the combustion chamber, typical values are in the order of a few bars; this pressure for staged combustion cycle engines has to exceed the chamber pressure by far, quite often by 2–3 times (see Table 3.3.8). The already extreme mechanical and dynamic loads require moderate turbine entry temperatures which should not exceed 900 K. In order to avoid local thermal overloading of the turbine

Table 3.3.8: *Characteristic values of gas generators.*

	Vulcain 2	SSME	LE-7	RD-0120	RD-180
T [K]	875	940/870	810	846	820
p_{GG} [MPa]	10.1	35/36	21.0	42.4	55.6
r_{of} [–]	0.9	0.89/0.8	0.55	0.81	54
$\dot{m}$ [kg/s]	9.7	80/30	53	78.6	887
P [MW]	5/14	56/21	4.5/19	62	93.5
p_c [MPa]	11.6	20.6	12.7	21.8	25.7

blades, a homogeneous temperature distribution is extremely important.

Table 3.3.8 summarizes key data such as chamber temperature and pressure, mixture ratio, propellant mass flow rate and power of the gas generator, and the appropriate main chamber pressure for several liquid propellant rocket engines. All except the Russian RD-180, which operates with LOX/kerosene, are H_2/O_2-fired engines, and while the European (Vulcain 2) is a gas generator cycle engine, the American (SSME), Japanese (LE-7) and Russian (RD-0120) engines are all based on the stage combustion cycle mode. While all the engines mentioned in Table 3.3.8 have a gas generator and turbine for each propellant, the Russian RD-0120 and RD-180 have only a single gas generator and turbine which drives both propellant pumps. All H_2/O_2 engines shown in Table 3.3.8 operate their gas generators in fuel-rich mode, independent of the engine cycle. By contrast, the RD-180, as in all Russian LOX/kerosene engines, operates in an oxygen-rich mode which is the outcome of a system-driven optimization of the engine. An analysis of equation 3.3.18 reveals that in fuel-rich H_2/O_2 systems, the smaller molecular mass of the driver gas results in a higher specific capacity c_p, which more than compensates for the disadvantage of the smaller mass flow rate $\dot{m}$ However, due to the comparatively large molecular mass of kerosene, gas generators which operate with LOX/kerosene in oxygen-rich operating mode yield smaller molecular masses of the driver gas. This has two advantages: high specific heat capacity c_p and much higher turbine mass flow rates. Additionally, oxygen-rich preburner operation apparently avoids the problem of soot formation and its successive deposition in the turbine or injection system.

The propellant injectors used in fuel-rich gas generators are often similar to those applied in the main chamber since the propellant temperatures are only slightly different. In order to establish stable combustion conditions in the case of oxygen-rich injection where the overall mixture ratio exceeds 50, a two-zone combustion process is realized. The first part of the oxygen is injected in the primary combustion zone to achieve gas temperatures of about 2000 K and the remainder of the oxygen is injected further downstream. Thus a homogeneous temperature profile is achieved at the entrance to the turbine. The problem of material compatibility of gas generator, fluid supply and turbine with hot, oxidizer-rich gases can be overcome by either surface passivation or appropriate coatings.

3.3.3.4 Turbomachinery

Liquid rocket engines can be differentiated into **pressure-fed** and **pump-fed** engines with respect to their **propellant feed system**. The tank mass of pressure-fed engines limits the chamber pressure. Pumps make high chamber pressures above 200 bar possible at moderate tank pressures (3–8 bar). The high chamber pressures required for lower stages necessitate pump-fed engines in order to limit the engine mass (and size) for large thrusts.

When designing **engines with turbopumps** it must be assured that the pressure at the pump inlet is above a minimum to avoid cavitation. The pumps, usually driven by one or more turbines, feed the propellants at the required flow rates and mixture ratios from the tanks via feed lines, valves, cooling channels and injection elements into the combustion chamber. The operational conditions are controlled by regulation devices.

Turbopump Components

The LOX pump of the Vulcain 2 engine is shown in Figure 3.3.28. The main components and systems of a turbopump power head are pump and turbine, gas generator or preburner, bearings, seals, gearbox, inlet and outlet manifolds, and lines. Only the most important components of turbopumps are described below. Detailed information on lubrication, bearings, seals and axial balancing systems can be found in the relevant literature. A turbopump consists of a **single or multistage pump** driven by a single or multistage turbine, according to the requirements. With few

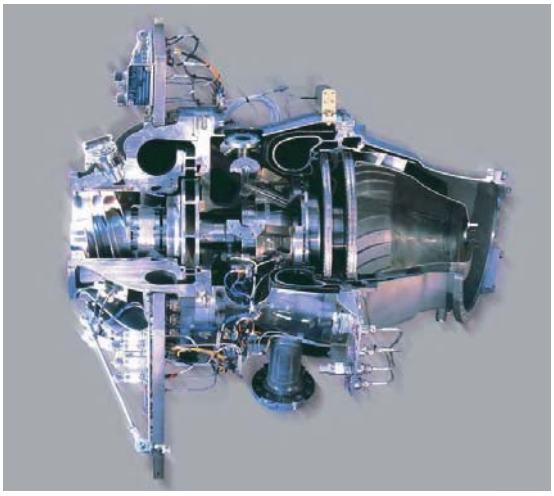


Figure. 3.3.28: *Vulcain 2 LOX turbo pump of the Ariane 5 (Source: ESA).*

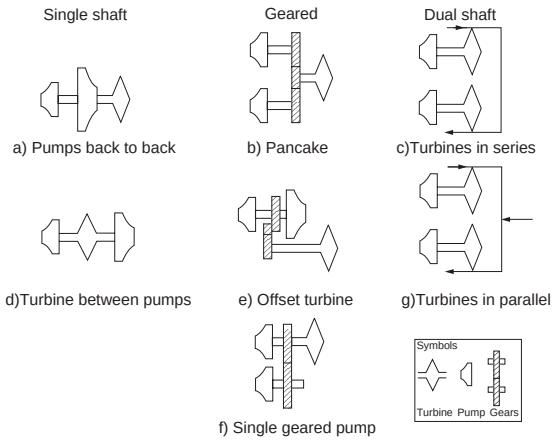


Figure 3.3.29: *Turbopump configurations.*

exceptions, only radial pumps are used in order to limit mass and size. Mostly axial turbines are used for the same reasons. With this approach tens of megawatts can be realized in compact turbopumps.

Different turbopump configurations are shown in Figure 3.3.29. The most compact type is the single-shaft turbopump. However, this concept requires a compromise concerning **pump speed**, which is identical for both pumps. To counteract this disadvantage the pump speed is adjusted by a gearbox. In a dual-shaft configuration the optimum speed and performance can be adapted individually for each propellant. A huge challenge for the design and construction of a turbopump is, besides controlling the **rotor dynamics**, the control of the high **temperature gradient**. Cryogenic fluids are often moved by the pump, whereas the turbine contains hot combustion gas.

Pumps

The main pump design parameters for incompressible steady-state conditions are as follows.

Increase of pressure per stage in Pa:

$$\Delta p = p_{\text{out}} - p_{\text{s}}$$

Enthalpy head in J/kg:

$$Y = \Delta p \,/\, \varphi = c_{u2}\, u_2 - c_{u1}\, u_1$$

Head rise in m:

$$H = Y \,/\, g$$

Tip velocity in m/s:

$$u = \pi\, d\, n \,/\, 60$$

Head coefficient [dimensionless]:

$$\Psi = 2\, Y \,/\, u^2$$

Flow coefficient [dimensionless]:

$$\varphi = c_{\text{m}} \,/\, u$$

Specific velocity per minute:

$$n_{\text{s}} = n\, \frac{\sqrt{\dot{V}}}{H^{0.75}}$$

Specific diameter in m:

$$d_{\text{s}} = d\, \frac{(H)^{0.25}}{\sqrt{\dot{V}}}$$

Suction specific velocity per minute:

$$n_{\text{ss}} = n\, \frac{\sqrt{\dot{V}}}{(NPSH_{\text{R}})^{0.75}}$$

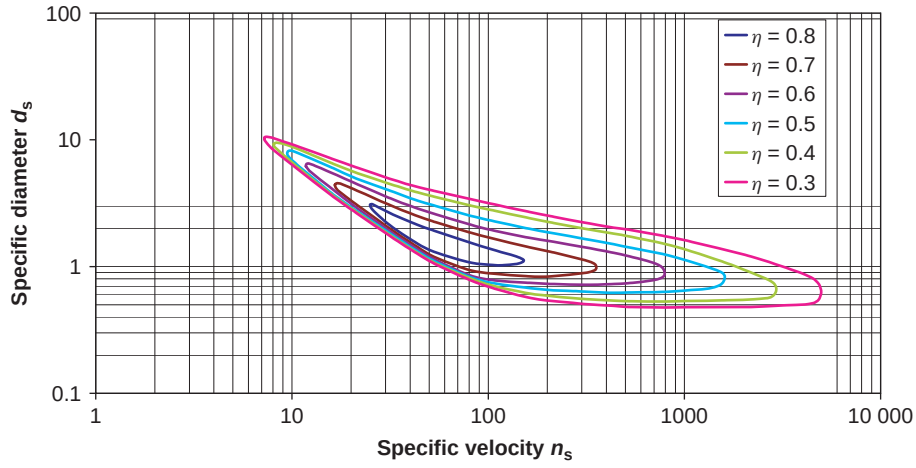


Figure 3.3.30: *Specific pump diagram n_s–d_s.*

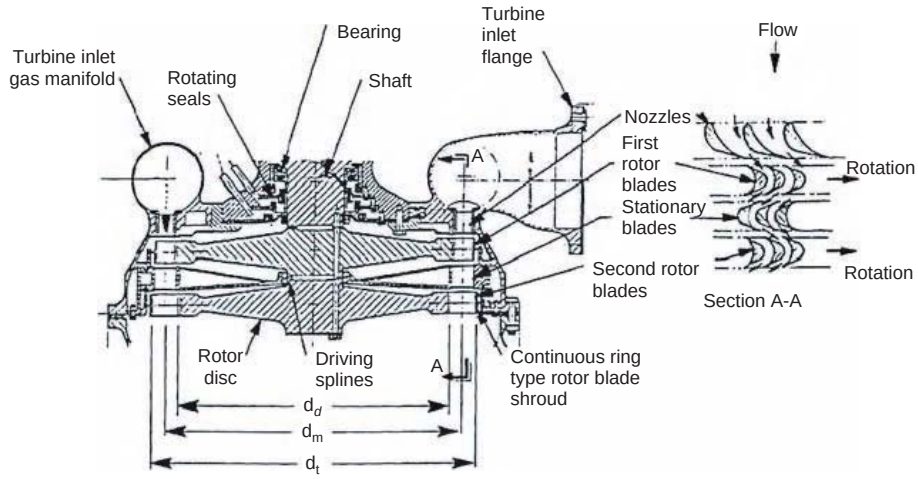


Figure 3.3.31: *Cross-section of an axial turbine (Source: Huzel).*

Net positive suction head in m:

$$NPSH_R = \lambda_c \frac{c_{m1}^2}{2g} + \lambda_w \frac{w_1^2}{2g}$$

Efficiency:

$$\eta = Y_{eff} / Y_{th}$$

Power in W:

$$p = \dot{m}\, Y / \eta$$

Head and flow rates as well as rotational speed are the characteristic parameters of a pump and determine the overall design and the rotor type. These parameters are related to the specific coefficients and to the real dimensions and operational parameters of a pump. The **efficiency** for the various design conditions is given in Figure 3.3.30.

Turbines

Axial turbines can be single stage or multistage (Figure 3.3.31). Each stage consists of a stator and a rotor. The stator with its fixed guiding vanes is located upstream, in which the fluid is expanded and accelerated. It is followed by the **rotor**, whose blades are moved by the force of the streaming gas. The kinetic energy is converted into mechanical energy. The two

main axial turbine types are impulse turbines and reaction turbines.

Impulse Turbine

In the impulse turbine the working gas is expanded in the stator and accelerated to high velocity. The acting momentum causes rotation of the rotor. The cross-section between the blades is constant and the gas flow is merely redirected. In the idealized friction-free case the gas is not accelerated, the pressure remains constant and the **reaction degree** is zero. The reaction degree characterizes the pressure drop in the rotor as a fraction of the pressure drop in the whole stage (stator plus rotor). The impulse turbine is used mainly in gas generator cycle engines with high pressure ratios and low flow rates; it exists in two design variants: **velocity-compounded** or **pressure-compounded**. In the first variant the gas is expanded only in the first stator; the flow is redirected only in the following rotor and subsequent stages. In the second variant the gas is expanded in the stator and is redirected in the rotor of every stage.

Reaction Turbine

In this turbine type the expansion is distributed between the stator and rotor. The cross-section between the blades diverges, causing an expansion. At a reaction degree of 100% the mechanical energy comes solely from the gas expansion in the rotor. Theoretically, any reaction degree can be designed, but mostly a design for a 50% reaction is used. The reaction turbine is used mostly in staged combustion engines to reduce the pressure ratio between the pump discharge and main combustion chamber.

Turbine Characteristics

The turbine efficiency is a function of the number of stages and the velocity ratio u/c_0, where u is the mean pitch line speed and c_0 is the isentropic pouting velocity, which is expressed for n stages as

$$\frac{u}{c_0} = \frac{\sqrt{\sum_1^n u_i^2}}{c_0} \tag{3.3.19}$$

Furthermore, the turbine efficiency depends on the type, whether reaction or impulse turbine. Figure 3.3.32 shows typical design points for a single-stage turbine. The curve does not consider losses caused by the gap between blades and the housing (**tip clearance**).

To account for tip clearance the values from the diagram have to be reduced to about 90–95%. As given by the characteristics, the efficiency increases with increasing velocity ratio u/c_0. How it relates to the allowable pitch line velocity of the blades, which is limited by the admissible temperature and blade root stress of the material, is given in Figure 3.3.33.

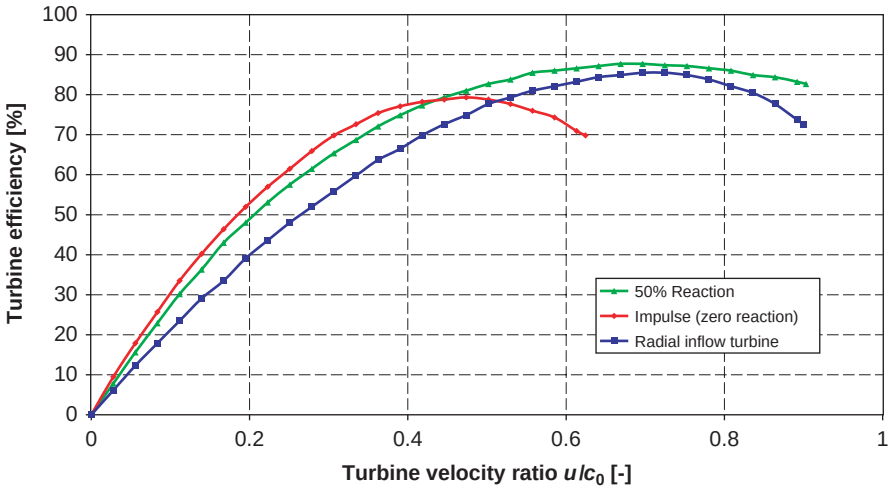


Figure 3.3.32: *Efficiency characteristics of a single-stage turbine.*

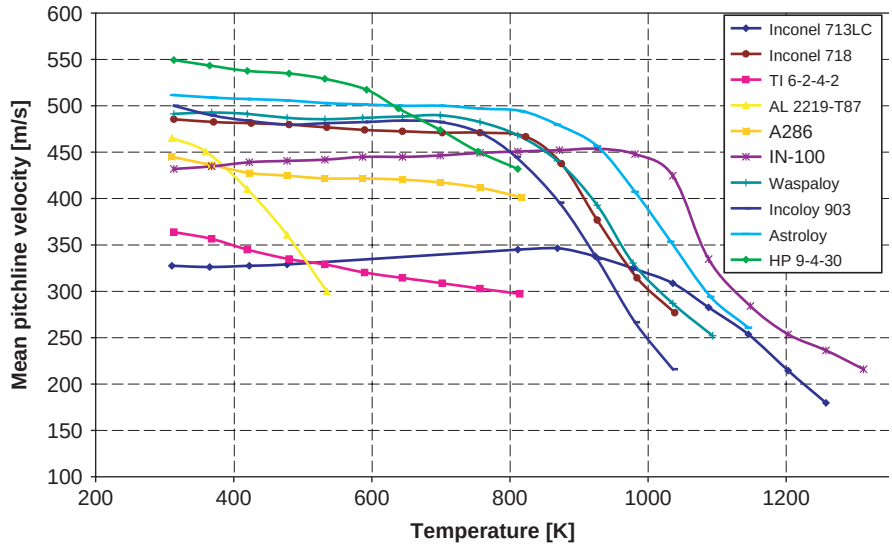


Figure 3.3.33: *Temperature dependence of turbine materials.*

3.3.4 Special Problems

Life of Thrust Chambers

The thrust chambers of rocket engines undergo high thermal and mechanical stress during operation, imposed by internal and external loads. Internal loads are the chamber pressure, the resulting thrust and the thermal loads along the hot gas wall and in the cooling channels. For example, temperatures from 20 to about 3800 K are reached during operation of the engine. Mechanical loads (aside from thrust loads) are induced by the inertial forces during gimbaling of the engine for thrust vector control and by transient side loads during engine ramp-up.

External loads are mainly pressure pulsations in the engine's ambient atmosphere, called buffeting, as well as the heating of the engine's outer surface by radiation from the solid booster flames.

These extreme thermal and mechanical loads can lead to two basically different kinds of failure: failure by **fatigue** and failure by **rupture**.

Repeated cyclic loadings, which are differentiated according to their frequency of occurrence into LCF (Low Cycle Fatigue) and HCF (High Cycle Fatigue) contribute to the fatigue, which is aggravated by the creep of the material under plastic strain over long operating periods (creep damage). The remaining free oxygen radicals generated by incomplete combustion degrade the life of the hot chamber wall further by oxidation (blanching).

The combustion chamber inner wall is the most stressed component of a thrust chamber and has the lowest expected lifetime. The accumulated damage is dominated by fatigue. The extreme temperature gradients between hot gas wall and cooling channel floor result in high plastic strains which ultimately lead to the so-called **dog house effect** during cyclic operations, that is the rupture of a cooling channel (see Figure 3.3.34).

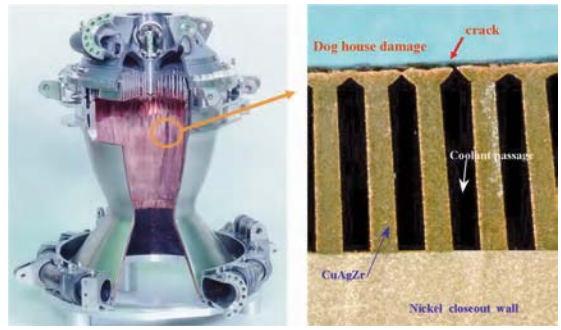


Figure 3.3.34: *Dog house effect (Source: ASTRIUM).*

Once opened, the crack is drastically cooled by the leaking coolant, which suppresses crack propagation. The risk of malfunction during operation is reduced by this effect, but is very high in the case of engine reignition.

A classic example for **sudden overstraining** is the occurrence of side loads during engine ramp-up. These side loads result from asymmetrical flow separation in the nozzle extension before the nozzle is fully flowing. These loads can reach up to 10% of the nominal thrust of the engine.

Oscillations in Propulsion Systems

As a consequence of its high energy density combined with small internal losses, the propulsion system of a launcher stage is a nearly **undamped system**. Therefore, special care has to be taken during the development process to design propulsion systems which avoid oscillations.

Oscillations occur either if there are interactions between the propellant tanks and the rocket engine, or if pressure oscillations are induced in the engine itself. The occurrence of such phenomena can lead to the loss of stages and consequently to the loss of the entire mission.

Three main types of oscillations can be distinguished, as follows.

POGO Oscillations

Oscillations induced by **resonances** between the engine, the structure of the stage and the propellants stored in the stage tanks are called POGO oscillations. Depending on the stage size, the frequency of POGO oscillations is between 80 and 100 Hz. POGO oscillations can be avoided by so-called anti-POGO devices built into in the feed lines between the tanks and the engine and acting as damping components.

Hydraulic Oscillations in the Propulsion System

This phenomenon is a low-frequency oscillation between the lines of the propellant feed system and the engine and is known as **chugging**. Frequencies are between 100 and 800 Hz.

It was pointed out in Section 3.3.3.1 (Injector Head) that a coupling between the liquid side of the feed system and the gas side of the combustion chamber can be avoided by designing a maximum possible injection pressure loss. Pressure-fed propulsion systems are much more vulnerable to chugging than are pump-fed systems because the turbopump already guarantees a high degree of decoupling.

Oscillations in the Combustion Chambers

Besides low-frequency oscillations, high-frequency oscillations are also triggered within the liquid propellant engine's combustion process. The triggering is induced by the coupling of the chemical reaction process with the eigenfrequency of the combustion volume. Resonances can occur within these oscillations in preferred frequency bands, which may lead to an **unstable combustion process**, called combustion instability. It results in oscillations with a very high energy content which may lead to the demolition of the combustion chamber. This dynamic combustion behavior has been the subject of intensive research for many years in an attempt to find an accurate mathematical formulation supported by comprehensive experimental investigations.

Such phenomena occur not only during the steady-state operational phase, but also during the **transients** while the engine is being ramped up or down. The associated sudden high-pressure fluctuations induce high accelerations and lead to the disintegration of the combustion jet boundary layer, which results in a prompt steep increase of the heat flux to the combustion chamber wall. High mechanical loads in combination with increased heat transfer to the chamber wall can cause the wall cooling system to fail in milliseconds, demolishing the engine because of the melting of the chamber wall and usually leading to loss of the mission as a consequence.

One of the most important issues during development of an engine is experimental proof of a **stable combustion process**. To increase operational reliability, damping devices such as acoustic cavities or symmetrically arranged baffles are additionally used.

In the Ariane program every engine used undergoes a hot acceptance test after its production to prove its dynamic stability.

It should be mentioned that the likelihood of combustion instability increases with the diameter of the combustion chamber.

3.3.5 Facilities for Rocket Engine Testing

The DLR research and test site in Lampoldshausen, Germany, is one of two European research and test centers where the ESA allocates testing activities for liquid rocket engine propulsion. The DLR-owned research and development facilities for small and subscale and component testing, namely the test stands P1, P2, P6 and P8, are located next to the ESA-owned facilities for full-scale components and engine tests. Figure 3.3.35 presents a bird's eye view of the testing site with the test facilities P1, P2, P3, P4, P5 and P8 and some of the storage areas for propellants and cooling water. The P1 test complex has facilities for vacuum simulation tests of satellite propulsion systems and apogee engines and the P2 complex is a test facility for sea-level tests of storable systems of less than 20 kN. The P3 stand is a full-scale thrust chamber test bench for cryogenic propellants with thrust range up to 2000 kN. The high-altitude simulation facility P4 for upper stage engines offers two test positions, one for storable and the other for cryogenic propellants. The cryogenic main stage engines of Ariane V, Vulcain 1 and Vulcain 2, were developed and tested at the P5 complex. The P6 facility has two test positions: P6.1 for green propellants (CH_4 and H_2), which went into operation in 2007, and P6.2 for cold gas nozzle testing. Finally, the P8 test stand is a European cryogenic rendezvous and docking facility which allows for subscale component tests at combustion chamber pressures up to 30 MPa.



Figure 3.3.35: *European Research and Test Centre at DLR Lampoldshausen (Source: DLR).*

During the development phase of an engine, all components of a propulsion system are usually tested in parallel and only later, using specific test facilities, are they qualified together as subsystems or as an entire engine. Such a development and testing methodology requires various large-scale facilities, which are a major cost factor. But the risk of a simple component failure occurring during a test, which damages major parts of a facility vital for engine development, thus endangering the program as a whole, is considered too high for other, less decentralized solutions.

System test facilities are characterized according to the engine types tested in sea-level and altitude simulation facilities. Booster and main stage engines are tested under sea-level conditions, upper stages, apogee and satellite engines under high-altitude conditions. Characteristic features of altitude simulation test benches are systems and installations which establish and maintain vacuum-like conditions during engine firing.

There is a general rule: fly as you test and test as you fly, and no component, subsystem or engine ever flies before it has been demonstrated that it fulfills all the requirements and operates within the safety margins. However, it is not always possible to realize this principle totally. For example, the ambient pressure during ascent decreases continuously, thus changing the thrust as well as the pressure difference across the thrust nozzle. It is extremely costly to modify a test bench for large rocket engines to simulate this ascent, and therefore such tests are often omitted. Additionally, installations typical of the launch site, such as the tower, the water cooling system or the operation of additional engines (large solid boosters), are nearly impossible to realize on a test stand.

P5, the cryogenic facility where all engines of the Vulcain family have been developed and qualified, is shown in Figure 3.3.36. Seen below are the green guiding tubes for the engine exhaust gases and immediately above the engine compartment with the red doors which are opened during testing. The liquid oxygen tank is installed in the tower behind the engine compartment to ensure turbopump entry conditions similar to the flight conditions. The liquid hydrogen tank is located behind the wall on the left and extends



Figure 3.3.36: *P5 engine test stand at Lampoldshausen (Source: DLR).*

below ground level for about two-thirds of its length. During engine firing, 2.5 m^3/s of water is necessary to ensure appropriate cooling of the exhaust gas guiding system and facility safety.

Some characteristic features of a typical altitude simulation facility are shown in Figure 3.3.37 where the new cryogenic upper stage engine Vinci is currently being tested (2008). The key components are the steam generator building at lower left where more than 250 kg/s of steam is produced during engine firing to maintain the desired operating conditions. Located immediately to its right is the condenser, which is connected by a large pipe to the test bench itself, shown in the upper part of this figure; it houses the engine, the diffuser and the propellant supply systems.

A schematic of the facility is shown in Figure 3.3.38. The length scale at lower left gives the dimensions of the bench. The propellant supply and the engine are located in the building to ensure as far as possible conditions similar to actual flight conditions. The engine



Figure 3.3.37: *Altitude simulation facility P4 with test positions P4.1 (left) and P4.2 (right) (Source: DLR).*

is fired directly at the water-cooled plug diffuser which is connected to the first ejector stage, which drastically reduces the temperature of the exhaust gases. The condenser, which is open to an underground reservoir, and the second ejector stage are the key components which guarantee the continuous cooling and repressurization of the exhaust gases. Since pumps are hardly able to handle these large flow rates efficiently, the exhaust gases are treated by applying the water jet pump principle. The hot, high-speed exhaust gases, more than 35 kg/s, from the engine require more than 250 kg/s of steam for repressurization. This steam is generated in five combustion chambers which operate slightly fuel rich with liquid oxygen and alcohol. The water mass flow rate needed to establish slightly overheated steam at about 2 MPa, together with the water necessary for cooling essential parts of the facility, amount to a total water mass flow rate of more than 4.5 m^3/s. In order to maintain the pressure level at the desired value, the water temperature in the system must not exceed 8 °C prior to testing.

3.3.6 Future Propulsion Systems

The issues described in the section on propulsion are strictly and intentionally dedicated to a description of today's flying systems and their components, as well as their design features and the tools used in their engineering. The existing chemical propulsion systems are currently working at their limits in terms of integrity

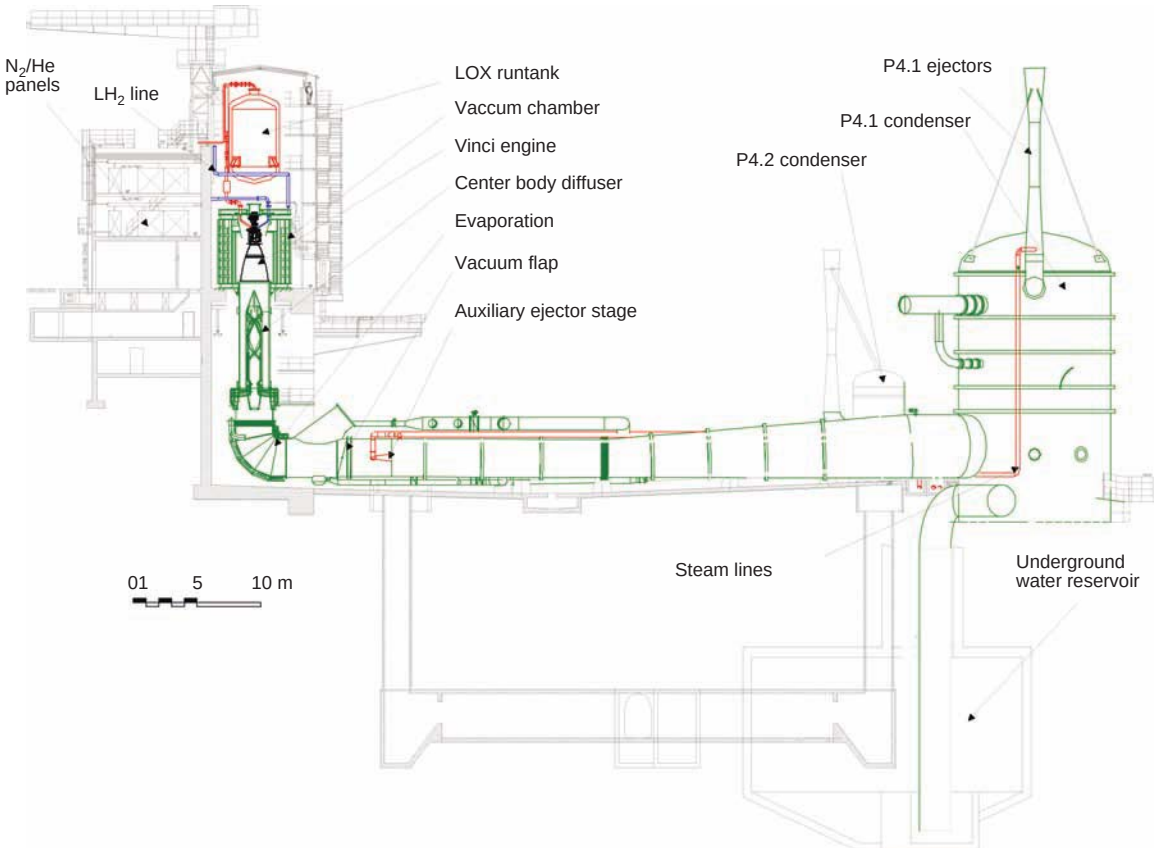


Figure 3.3.38: *Schematic of the high-altitude simulation facility P4.1 (Source: DLR).*

and performance. It can therefore be stated that any further performance increase will be marginal. The limiting components are mainly the turbomachinery and the combustion chamber because of the enormous thermal, mechanical, static and dynamic loads to which they are subjected. Even assuming that a significant technology step could be taken toward the development of very light materials with high thermal resistance and low thermal expansion coefficients, it would only result in a marginal performance increase. In looking for concrete improvement options, the following steps seem reasonable:

- Improvement of cooling methods, to result in lower pressure losses.
- Increase of turbine inlet temperatures and rotational speed.
- Management of material compatibility for preburners operating at high mixture ratios.

As described in Section 3.3.3.2, the next logical step in increasing launcher performance is to find engineering solutions for nozzles capable of adapting to ambient pressure during ascent.

From the very beginning of commercial space flight, analytical as well as experimental investigations have been performed on the subscale level with the goal of using oxygen from the surrounding air during the flight through the atmosphere in order to increase the specific impulse and thereby substantially decrease the take-off mass of space vehicles. Recent investigations by NASA have concluded that a substantial benefit from such systems would only be visible if

the separation Mach number between the first and upper stage could be increased to the range of 8–12. This requires operation of the combustion chamber at Mach numbers in the range of 4–8. A propulsion system operating under these conditions to provide a positive thrust could not even be realized today under laboratory conditions. As a conclusion one can say that the next launcher systems to be developed will remain with state-of-the-art chemical propulsion.

Bibliography

[3.3.1] Huzel, D.K., Huang, D.H. *Modern Engineering for Design of Liquid-Propellant Rocket Engines*, Progress in Astronautics and Aeronautics, Vol. 147. Reston, VA: AIAA, 1992.

[3.3.2] Yang, V., Anderson, W. (eds.) *Liquid Rocket Engine Combustion Instability*, Progress in Astronautics and Aeronautics, Vol. 169. Reston, VA: AIAA, 1995.

[3.3.3] Kuo, K. (ed.) *Recent Advances in Spray Combustion: Spray Atomization and Droplet Burning Phenomena, Volume 1*, Progress in Astronautics and Aeronautics, Vol. 166. Reston, VA: AIAA, 1996.

[3.3.4] Kuo, K. (ed.) *Recent Advances in Spray Combustion: Spray Combustion Measurements and Model Simulation, Volume 2*, Progress in Astronautics and Aeronautics, Vol. 171. Reston, VA: AIAA, 1996.

[3.3.5] Sutton, G.P. *History of Liquid Propellant Rocket Engines*. Reston, VA: AIAA, 2006.

[3.3.6] Isakovic, S.J., Hopkins Jr., J.P., Hopkins, J.B. *International Reference Guide to Space Launch Systems*, Third Edition. Reston, VA: AIAA, 1999.

[3.3.7] Sutton, G.P. *Rocket Propulsion Elements*, Third Edition. New York: John Wiley & Sons, Inc., 1963.

[3.3.8] NASA Facts. Next Generation Propulsion Technology: Integrated Powerhead Demonstrator, Technology. *FS-2005-01-05-MSFC*, Publ. 8-40355, 2005.

[3.3.9] Cai, C., Jin, P., Yang, L. *et al.* Experimental and Numerical Investigation of Gas-Gas Injectors for Full Flow Stage Combustion Cycle Engine. *AIAA-2005-3745*, 2005.

[3.3.10] Davis, J.A., Campell, R.L. Advantages of a Full-Flow Staged Combustion Cycle Engine System. *AIAA-1997-3318*, 1997.

[3.3.11] Wade, M. Encyclopedia Astronautica. www.astronautix.com, 1997–2008.

[3.3.12] Cannon, J.L. Turbomachinery for Liquid Rocket Engines. Liquid Propulsion Systems – Evolution and Advancements. AIAA Professional Development Short Course, 2003.

[3.3.13] Liang, K., Yang, B., Zhang, Z. Investigation of Heat Transfer and Cooling Characteristics of Hydrocarbon Fuels. *J. Propul. Power*, **14** (5), 1998.

[3.3.14] Hammond, W. *Space Transportation: A System Approach to Analysis and Design*. Reston, VA: AIAA, 1999.

[3.3.15] Newell, J.F., Rajagopal, K.R. Probabilistic Methodology – A Design Tool for the Future. www.engineeringatboeing.com/articles/probabalistic.htm, 1999.

[3.3.16] May, L., Burkhardt, W.M. Transpiration Cooled Throat for Hydrocarbon Rocket Engines. *NASA KEE6-FR*, 1991.

[3.3.17] Meinert, J., Huhn, J., Serbest, E. *et al.* Turbulent Boundary Layers with Foreign Gas Transpiration. *J. Spacecr. Rockets*, **38** (2), 191–198, 2001.

[3.3.18] Keener, D., Lenertz, J., Bowerson, R. *et al.* Transpiration Cooling Effects on Nozzle Heat Transfer and Performance. *J. Spacecr. Rockets*, **32** (6), 981–985, 1995.

[3.3.19] Hagemann, G., Immich, H., Nguyen, T. *et al.* Advanced Rocket Nozzles. *J. Propul. Power*, **14** (5), 620–634, 1998.

[3.3.20] Korte, J.J., Salas, A.O., Dun, H.J. *et al.* Multidisciplinary Approach to Aerospike Nozzle Design. *NASA TM 110326*, 1997.

[3.3.21] Frey, M., Hagemann, G. Critical Assessment of Dual-Bell Nozzles. *J. Propul. Power*, **15** (1), 137–143, 1999.

[3.3.22] Haidn, O.J., Greuel, D., Herbertz, A. *et al.* Application of Fiber Reinforced C/C Ceramic Structures in Liquid Rocket Engines. In Assovskiy, I.G., Haidn, O.J. (eds.), *Space Challenge in XXI Century*. Moscow: Torus Press, pp. 46–72, 2005.

[3.3.23] Sebald, T., Beyer, S., Gawlitza, P. *et al.* Advanced Thermal Barrier Coatings for High Heat Fluxes in Thrust Chambers of Liquid Rocket Engines. 4th International Conference on Launcher Technology Space Launcher Propulsion, Liège, 2002.

[3.3.24] Schulz, U., Fritscher, K., Peters, M. *et al.* Fabrication of TBC-armoured Rocket Combustion Chambers by EB-PVD Methods and TLP. *Sci. Technol. Adv. Mat.*, **6**, 101–110, 2005.

[3.3.25] Stark, R., Haidn, O.J., Böhm, C. *et al.* Cold Flow Testing of Dual-Bell Nozzles in Altitude Simulation Chambers. Proceedings of EUCASS European Conference for Aerospace Sciences, Moscow, 2005.

[3.3.26] Kau, H.P., Langel, G. Raumfahrtantriebe. Vorlesung TU München Lehrstuhl für Flugantriebe, Sommersemester, 2007.

[3.3.27] Schmidt, G. *Technik der Flüssigkeits-Raketenantriebe*. DaimlerChrysler Aerospace, 1999.

3.4 Launch Infrastructure

Dirk Rüdeger Albat

An overview is given of the infrastructure of a launch complex, the related support equipment, the preparation of the launch vehicle and the payloads for the launch, and other peripheral equipment needed at a launch site.

3.4.1 Requirements and Missions

Launch infrastructures have to comply with a broad set of requirements. Above all, launch complexes have to be operated **without risks** for people and goods. It is therefore important that the launch site provides sufficient free space around hazardous areas. The launch trajectories must be outside of populated areas and a broad range of orbits should be economically reachable. Therefore, launch sites near the equator offer clear advantages for geostationary orbits, as nearly no correctional maneuvers are needed for inclination changes and the extra energy provided by the Earth's rotation can be fully exploited. The topology of the launch site should allow the construction of the required infrastructure. Nearby hills help, for example, to achieve the optimal positioning of antennas needed to receive the telemetry stream emitted from the launch vehicle or for tracking the launcher during its ascent.

The available infrastructure around the launch site should allow affordable transport of the launch vehicle elements, the payloads, the launch teams, the propellants, and all other types of ground support from the various production sites to the launch site.

The **climate** at the site should allow regular launches. Sites with limited and predictable probability of thunderstorms or high-altitude shear winds hold the promise of high availability rates. The risks of tornados at the launch site should be minimal, and the site should preferably not be located in a region with earthquake risk.

Launch complexes require substantial budgets to finance construction and maintenance. Predictable economic and **political stability** of the region around the launch site helps to secure the required finance.

The **infrastructure of a launch site** makes possible the following tasks:

- Arrival, unloading and transport of the launch vehicle elements, the payloads and all other equipment and materials not produced in situ.
- Assembly and test of the launch vehicle and the payloads.
- Loading of the launch vehicle and the payloads with propellants and gases and providing charge capabilities for the on-board batteries.
- Assembly of the payloads on the launch vehicle.
- Transport of the assembled launch vehicle including the payloads to the launch pad.
- Accessibility for the last preparatory activities before the launch, such as cryogenic propellant fueling or switching on the on-board systems.

3.4.2 Concepts

The design of a launch complex has to take various criteria into account. Commercial launch sites often prepare several payloads and launch vehicles in parallel. This requires adequate coordination: delays either of the launcher or of payloads should have minimum collateral impact. The consequences of a launch failure should remain predictable and the repair costs of and recovery schedules from damage should be manageable. The maximum allowable preparation cycles for launchers and payloads must be defined to accommodate the specified launch rates.

In the last 50 years the following concepts for launch complexes have been realized.

The **"all-in-one" concept:** Launch vehicle assembly, test and launch preparation are performed directly on the launch pad on the launch table. This allows minimized budgets and lead times for construction of the ground infrastructure, but launch rates are also limited (a typical value is five launches per year). Alternative processing options in case of delays do not exist, and in case of a launch failure in the early ascent phase, there is a very high risk that the overall launch preparation infrastructure will be damaged.

This concept was applied mostly in the 1950s to 1970s, in Europe for the Ariane versions 1 to 3, for example. The application of the all-in-one concept is today limited mostly to small launch vehicles. The next logical evolutionary step was the **physical separation** between launch vehicle preparation and launch activities; the disadvantages of the all-in-one concept led in the 1960s and 1970s to concepts where the launch pad is used only for activities relevant for the launch proper. Launch vehicle and payload assembly and tests are performed in separate buildings before the launcher is transferred to the pad. This concept was applied in the USA for Saturn V and the Space Shuttle, and in Europe for Ariane 4.

The **"clean pad" concept** (applied from the 1990s onwards, e.g., for Ariane 5) aims at minimizing the amount of infrastructure needed on the launch pad. An important part of the infrastructure is incorporated directly in the launch table and only the absolute minimum of infrastructure remains on the pad. But this concept also has one major disadvantage: even minor anomalies on the launch vehicle may require a transfer back from the pad to the assembly building, as the clean pad concept does not in many cases provide appropriate access for repairs.

3.4.3 One Realized Example: Ariane 5

This section describes the infrastructure required for an Ariane 5 launch.

Payload preparation is performed in **clean rooms** (see Figure 3.4.1) equipped with mechanical and electrical equipment. After the (in general, aerial) transport from the manufacturer to the launch site, the payload is unloaded and connected to its specific ground support equipment, allowing the testing of all electrical and mechanical functions of the subsystems. After final leak checks and functional controls, the propellant tanks of the spacecraft are filled and high-pressure vessels are pressurized with helium or nitrogen. The customer teams themselves carry out this test and preparation phase, with the launch provider giving logistical support (e.g., means of transport, climate control equipment, electrical supply, workshops, offices, etc.). At the end of this preparatory



Figure 3.4.1: *Payload preparation (Source: ESA).*

phase the payload is weighed to make sure that the allowable mass budgets are met. Then the payload is handed over to the launch provider for integration on the launch vehicle. The overall preparation cycles for payloads vary from four to eight weeks and might even be longer for science payloads.

The preparation of the launch vehicle starts in the **vehicle assembly building**. The different stages of the launcher are extracted from their transport containers, erected with the help of a crane, assembled, and connected to the fluid and electrical ground support equipment. The assembly of the launch vehicle is followed by an intense test phase, including leak checks of all fluid systems and engines and the functional testing of all systems.

For a heavy launcher of the Ariane 5 class this part of the launch campaign requires about 20 working days.

After the end of the integration and test phase the launcher and the launch table are transferred to the final assembly building, as shown in Figure 3.4.2.

Payload integration on the launcher: The launcher and payloads remain in the final assembly building for about eight days. The following working steps are performed in this building:

- Assembly of the lower payload including its payload adapter directly on the launcher.
- Off-line assembly in a separate preparation workshop of the upper payload, the launch vehicle carrying structures and the launcher fairing.



Figure 3.4.2: *Ariane 5 assembly building (Source: ESA).*



Figure 3.4.3: *Assembly of the payloads on the launcher (Source: ESA).*

- Transport of this upper composite into the final assembly building (which has a height of 90m) and mounting on the launcher (see Figure 3.4.3).
- Testing of all electrical and mechanical connections.
- Dress rehearsal including a simulated countdown and flight to demonstrate that the complete launch infrastructure (launcher, payloads, control centers, tracking stations, ground station network for telemetry) is well coordinated and working together properly.
- Filling of the upper stage (in the case of storable propellants, such as for the EPS upper stage).

The launcher is now ready for rollout and transfer to the launch pad, which is about 3 kilometers away from the final assembly building.

The next step is the propellant filling of the cryogenic stages on the launch pad, followed by the final countdown and the launch. The pad consists of the following elements (Figure 3.4.4):

- Exhaust ducts (two ducts for the solid propellant boosters and a third duct for the main stage liquid propulsion engine).
- A water tower 90 m in height allowing a high water flow rate; just before the launch, water fountains are directed onto the launch table to reduce the acoustic loads during liftoff.
- Electrical and fluid lines to link the launcher to ground tanks and to the launcher control center.

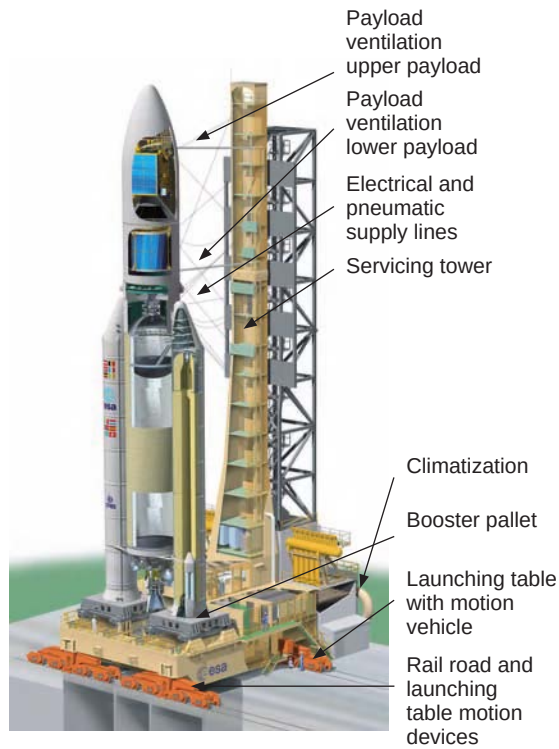


Figure 3.4.4: *Ariane 5 launch table (Source: ESA).*

Launcher control center and countdown: The launcher is transported to the pad about 12 hours before liftoff. The operations needed for launch preparation last about 8 hours and are directed by the launcher control center (see Figure 3.4.5), about 3 kilometers away. The launcher control center is equipped with its own fully autonomous oxygen supply and independent electric generators for security reasons. The walls and roofs of the center are reinforced like a bunker. All launcher activities are directed from this center, as follows:

- T0 − 6 h: Start of the **countdown**. Preparation of the ground facilities for filling the ground propellant lines, evacuation of personnel from the launch pad.
- T0 − 5 h: Start of cryogenic main stage propellant filling.
- T0 − 4 h: Start of cryogenic upper stage propellant filling.
- T0 − 1 h: Release of mechanical safety barriers on the launcher's pyrotechnical systems.
- T0 − 6 min: Start of the automatic launch sequence.

This sequence is controlled in a fully automatic mode by two redundant computer systems and includes the following tasks:

- Start of the on-board flight program, initialization of navigation, guidance and control systems.
- Pressurization of the cryogenic propellant tanks to flight level.
- Isolation of the launcher from the ground support lines, downstream leak checks behind the launcher's isolation valves.
- Chilling down of the main engine's combustion chamber with on-board propellants.
- Start of the water stream on the launch table.
- Release and retraction of the upper stage fueling arms.

At T0 − 4 s the launcher's on-board computer takes over the management of the last controls, starts the Vulcain main engine, analyzes engine health status after start, and ignites the solid boosters which initiate the liftoff.

The Launch Control Center

While the activities in the launcher control center during the day of launch resemble those of pilots who work through a checklist in their cockpit, the launch control center (see Figure 3.4.6) plays more the role of an airport tower.

The launch control center assures the **coordination of all needed subsystems**, as follows:

- Telemetry reception allowing real-time on-board data transmission during the flight.
- Launcher localization and tracking of the flight path.
- Flight safety.
- Ground energy supply.
- Synchronization of time signals coming from or provided to all locations of the activated network.
- Video transmission.



Figure 3.4.5: *Ariane 5 launcher control center (Source: ESA).*



Figure 3.4.6: *Ariane 5 launch control center (Source: ESA).*

- Data transmission and communication between the participating centers.
- Payload surveillance and coordination.
- Command desk of the mission director.
- Weather station.

A system of **connecting roads** allows transport of flight hardware. The launcher and its launch table are moved on rails between the preparation buildings and the launch pad. The complete compound, consisting of the launcher, the launch table and mobile air cooling units (needed for payload conditioning under the fairing), is moved by truck (Figure 3.4.7). The transport speed is about 3 km/h.

On-board fluids (such as liquid helium and nitrogen, water, propellants, etc.) are partly produced on the launch site and are stored in ground tanks (Figure 3.4.8).



Figure 3.4.7: *Ground transport of Ariane 5 (Source: ESA).*



Figure 3.4.8: *Production plant for liquid hydrogen for Ariane 5 (Source: ESA).*

A **weather station** (Figure 3.4.9) allows evaluation of the weather conditions for all transport and for the launch day. The authorization for transport or for the launch itself is linked to predefined criteria, such as maximum allowable wind speeds or thunderstorm probabilities.

Among others, the following measurements help to establish the diagnosis:

- Ground wind measurements are made at altitudes between 0 and 100 m via wind masts.
- Doppler sonars allow the measurement of wind speeds at altitudes in the range of 100–500 m. This is accomplished by emitting different frequencies and measuring the frequency shifts of the reflected waves, which permits computation of the corresponding wind speed.
- Altitude winds in the range from 500 to 10 000 m are monitored via laser or balloons in order to detect dangerous shear winds.
- Radar measurements allow continuous observation of cloud thickness and precipitation intensity, and considerably improve the forecast quality for estimating thunderstorm probability. If any doubts remain, weather balloons may be released just before liftoff to confirm the ground measurements.

After liftoff, **tracking stations** (Figure 3.4.10) allow observation and real-time analysis of the flight path. From the beginning of the flight until the end of the launcher mission (payload separation), launcher and



Figure 3.4.9: *Weather station (Source: ESA).*



Figure 3.4.10: *Tracking station (Source: ESA).*

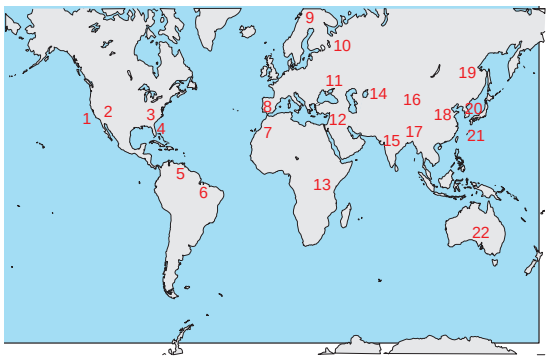


Figure 3.4.11: *Worldwide launch sites (Source: Space Today online).*

payloads are under permanent surveillance. This requires telemetry antennas, optical means, radar stations, and infrared and optical cameras.

Flight trajectory and behavior are continuously evaluated so that potential dangers are identified immediately. In extreme cases the launcher can be destroyed by remote command to exclude all risks for populated areas.

Table 3.4.1: *Geographical coordinates of launch sites.*

Launch site	Coordinates	Launch system
Kourou (French Guiana)	5°N, 53°W	Ariane 5, Vega, Soyuz
Cape Canaveral (USA)	29°N, 81°W	STS, Titan, Atlas
Vandenberg (USA)	35°N, 121°W	Titan, Atlas, Delta
Baikonur (Russia)	46°N, 63°E	Proton, Soyuz, Zenit
Plesetsk (Russia)	63°N, 41°E	Soyuz, Cosmos, Rockot
Sriharikota (India)	14°N, 80°E	PSLV, GSLV
Jiuquan (China)	41°N, 100°E	CZ (Long March)
Tanegashima (Japan)	30°N, 131°E	H-II, H-IIA

3.4.4 Major Launch Sites

Over 20 launch sites have been built worldwide (see Figure 3.4.11). Table 3.4.1 lists the most important ones with their geographic coordinates and the systems launched there. The special case of the Odyssey launch site should also be mentioned. It is a floating ocean platform allowing the Russian Zenit rocket to be launched from various locations.

3.5 System Qualification

Jörg Krüger and Anton Grillenbeck

3.5.1 Introduction

Development Phases of Launch Vehicles

The starting point for the development of any launch vehicle is understanding and considering the customer's needs and primary objectives, usually defined in a statement of work and a functional requirements specification. Generally the most important parameters are:

- System reliability
- Payload capacity and accuracy for target orbits/missions

- Budget restraints for the development phase
- Costs for production and operation
- Date of the first flight (maiden flight), launch sequence and required production rate.

Often, the development of launch systems follows the principles of **total quality management** (TQM), meaning that system reliability and robustness are of main concern, and thus the design drivers. However, if other parameters are considered to be of equal importance, the development approach is called **design to value**, or **design to cost** in the particular case when development costs are faced with production and operating costs during an optimization process.

In addition, launch vehicle development programs for civil applications are often combined with other programmatic aspects such as:

- To enable the participation of companies according to predefined shares among different countries (so-called GEO return within the ESA perimeter).
- To use explicitly new technologies which have been funded beforehand, such as propulsion technologies, materials, or other specific stage subsystems.
- To ensure the system's availability at a given time; for example, to put a planned satellite system into orbit on time.
- To use a certain launch site.
- To use already available components or technological heritage (from past projects).
- To respect general applicable standards (e.g. debris avoidance).

The main customer for civil launch vehicle development in Europe is ESA, assisted by the French Space Agency CNES. ESA/CNES provide the technical requirements via a **functional requirements specification** (FS, cf. ECSS–E–10 Part 6A, in the Ariane context called CdCF (*Cahier de Charge Fonctionnel*)). The launch vehicle developer is in charge of formally responding to the functional specification by implementation of a **technical specification** (TS) which describes in detail the desired performance and the functional breakdown of the intended product derived from the given main functions. The functional description, usually set up as a tree of functions, is the basis for later elaborating the technical solution, detailed design and systematic verification approach for each required and specified function from the highest to the lowest system level; see Figure 3.5.1 for an example of a function tree. Drawing up a functional analysis loop makes it possible to analyze all functions for the entire life cycle of the product. It is essential to identify all relevant loads, interfaces, and interactions.

In a second step, the identified loads, the specific requirements and interface conditions are quantified and in turn laid down in the technical specifications, in general dimensioning loads, or in interface control documents (ICDs).

The appropriately detailed quantification of all requirements and loads on the system and main subsystem levels is the key challenge in the early phase of the launch vehicle development project, besides the elaboration of design and operational trade off studies. During launcher development, this process has a strongly vertical dimension: the loads and requirements are detailed from the launch vehicle level to the stage level and down to the subsystem level. According to the so called V-Model, the follow-up process at each level consists of: requirements definition, design and dimensioning, manufacturing, test and verification. Toward the end of the development process, starting with the lowest subsystem level element and then moving up to each next higher level, the integrity and functional performance of the product has to be proven by verification to the technical specification. The necessary process and tasks for this qualification process are the main focus of this chapter.

System engineering plays a central role in the launch vehicle development process, since it is responsible both for system definition and for organizing the means of verification as an integral part of the overall qualification process, commencing with the concept phase.

The role of system engineering is manifold: besides the management of the margins projectwide, another main task is the elaboration of the core product development documentation, consisting of (the designations used in the Ariane program are given in brackets for reference):

- Specification of system performance (ST, Technical Specification)

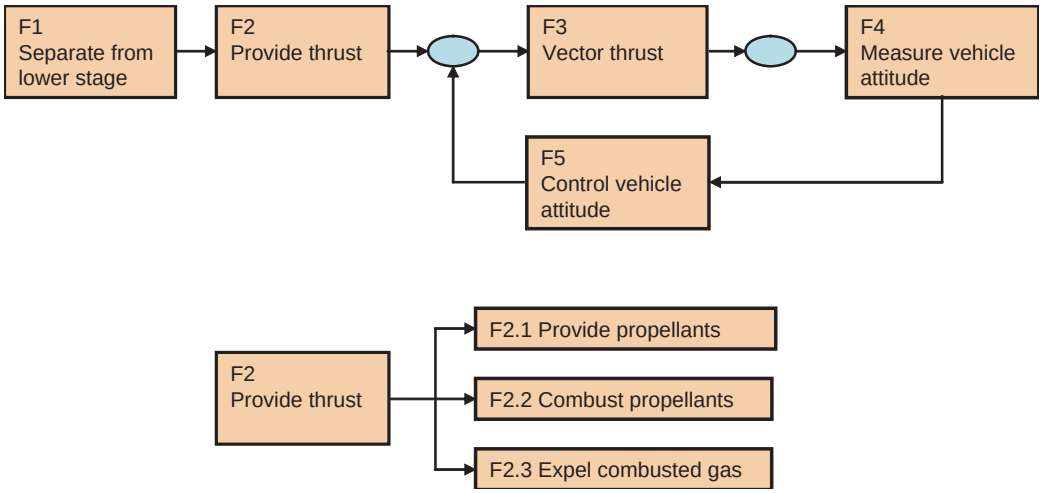


Figure 3.5.1: *Sample function tree describing the operation of a product to be developed.*

- Definition of the overall system (DF, Functional Definition File)
- Specification of system interfaces (SI, Interface Control Document)
- Trade-off analysis, selection of preferred solutions and justification (DJ, Justification File)
- Identification of critical elements (Critical Point List)
- Definition of the verification philosophy and methods (PVE, Requirement Verification Plan)
- Definition of a measurement plan for in-flight qualification and flight data analysis (FAP, Flight Assessment Plan).
- User manuals.

System engineering provides the link between all the required technical disciplines involved in the launcher development since it is responsible for global load characterization, materials selection, system design, and the integration of components into the overall system. The realistic characterization of the product under the given loads (environmental, induced and operational) is the most challenging effort during the later qualification process due to the large variation and superposition of combined loads.

Technical Qualification of the Launch Vehicle

The qualification approach is different for the various technical disciplines. For example, for the design of the **mechanical systems**, narrow margins of safety are used to achieve low structural mass in order to allow a high payload lift-up capacity while complying with the required reliability standard. Therefore, starting with the processing of the raw materials, all production steps have to provide certified and constant mechanical parameters subject to comprehensive characterization tests. Particular attention is paid to the assessment of **fatigue failures** and the prediction of crack propagation. To this end, load cycles and dynamic load assumptions are defined conservatively. In contrast, thermal engineering applies a different approach for the thermal analysis and qualification of the thermal design. Usually the worst case is assumed, where the most unfavorable combination of thermal system conditions must not endanger any of the functions and operations.

Often during the design and analysis phase the assessment and calculation cycles have to be repeated iteratively in order to consider changes in the design, operations or loads. Only robust and continuous **management of margins and risks** enables the convergence of design and requirement on the one hand, and the optimization of performance versus development costs on the other hand. For instance, increasing system mass might be balanced by margins in the velocity increment of the first stage, or by the specific impulse of the upper stage.

Functional and Structural Verification

The functional and structural verification of the launch system ensures that the product complies in all details with its **specification** and that all **safety margins** are positive. This mainly comprises the proof that the required performance is delivered under realistic environmental and operational conditions, that all requirements are fulfilled, that there are no unexpected deviations in the product hardware and software, and that the ground segment launch of operations have been formally approved.

Most of the **verification measures** are already implemented by consequent and comprehensive application of a formally qualified development process (e.g., analysis of all requirements for the complete life cycle, analysis of loads, etc., up to the proving methods). Typical verification methods for launch systems are numerical analysis and simulation, experimental demonstration, analogy, inspection, and test or qualification flight. Usually a combination of these methods is applied. In particular, analyses and tests are performed stepwise and complementarily on all product levels during the development process. Simulation and analysis models have to be validated. Critical assumptions and hypotheses have to be proven by experiments/tests or statistical approaches.

Verification by analogy is often used in the context of analysis. Here, it is shown that a certain part will behave similarly to another part which has already been qualified for another project under equivalent or even more severe requirements or load conditions. This, of course, requires that all information concerning the previous qualification process is available, including the exact hardware configuration and the applied operational conditions. If not, further delta qualification steps might be required.

Experimental demonstration is performed with the actual hardware or software in a simulator or in a simulated operational environment. In the framework of the Ariane program, further demonstration tests are performed within the ARTA program after launcher qualification for statistical proof of low deviations and tolerances. In this program, parts from the actual manufacturing process are taken and subjected to representative investigations.

Typical **inspection methods** are nondestructive, like visual inspections, X-ray investigation, weighing, helium leak tests, etc. Others involve regular monitoring of the manufacturing processes and the respective facilities.

Acceptance tests (see Table 3.5.1) are performed regularly at all system levels (components, subsystems and the entire system). In contrast to the more detailed and comprehensive development and qualification tests, the acceptance tests are usually performed in the later production and operation phase in order to limit risks. Acceptance tests provide the proof that a part or system is within the specified tolerance concerning its functions and that it is free of defects and integration failures. The test loads applied during the acceptance test are significantly below the dimensioning loads applied in the qualification phase. Acceptance test conditions are also applied after refurbishing the launcher or when replacing components after a launch abort.

Usually **qualification tests** are distinguished as being either mechanical or functional qualification tests, and ground or flight qualification tests. Ground qualification tests might be performed on prototype models, proto-flight models or scaled models. Qualification tests may be limited to certain dimensioning flight phases or cover the full extent of the life cycle. These tests may be designed for a specific load case or for a combination of different load cases. Some tests may be performed until the failure of the unit under test, others simply within the nominal operational conditions for the unit. This variety is driven by the need to cover all critical areas of the product life cycle. Sometimes it is difficult to deduce the reliability of a system from a single qualification test, even though the test was performed until rupture.

It is practically impossible to simulate all **flight conditions** in a single test facility, particularly when a combination of conditions or loads has to be considered simultaneously. In addition, sometimes not all parameters are accessible by direct measurement so they have to be determined via other representative measurements. In some cases the increased intensity of other test parameters, for example the number of load cycles, may help to deduce the test objective as well.

Qualification tests are used to consolidate the required margins of safety and to characterize operational and functional behavior. These tests also enable

Table 3.5.1: *Typical acceptance tests.*

Potential failure mechanism	Primary acceptance tests to precipitate failure mechanism									
	Functional	Wear-in	Vibration or acoustic	Shock	Thermal cycle	Thermal vacuum	Leakage	Proof pressure	Proof load	EMC
Parameter drift	x		x		x	x				
Electrical intermittants			x	x	x	x				x
• Solder joints										x
• Loose wires										x
• Connectors										x
Latent defective parts	x		x	x	x	x				
Parts shorting			x							
Chafed/pinched wires			x							x
Adjacent circuit board contact			x	x						
Parameters changing due to deflections			x		x	x				x
Loose hardware			x	x						x
Moving parts binding		x				x				
Leaky gaskets/seals					x	x	x			x
Lubricants changing characteristics		x			x	x				
Material embrittlement				x	x	x				
Outgassing/contamination			x	x	x	x				
Degradation of electrical or thermal insulation						x				x
Corona discharge/arcing						x				x
Defective pressure vessels								x		
Structural defects									x	
Defective wiring	x									x
Defective tubing							x			

the verification or calibration of analytical models which may be used later for flight prediction.

Typical dimensioning **load cases for qualification tests** (see Table 3.5.2) are the determination of structural or functional behavior under static loads, under high acceleration, high vibration and acoustic loads, mechanical or thermal shocks, pressure gradients, vacuum environment, aerodynamic flutter, and impact of particles or radiation. Further details are discussed in Section 3.5.3.

In case of an unexpected failure during qualification tests, a general rule is to solve the problem at the **lowest possible integrated level**; that is, at the unit level rather than at the component level, at the subsystem level rather than at the stage level. If a failure occurs on a higher level in the verification process it often results in time- and budget-consuming investigations.

Launch Readiness

The clearance of a launch vehicle for its maiden flight lift-off has several prerequisites. The **verification** process must be concluded, with the exception of those items which need the maiden flight for qualification and verification purposes. The **launch vehicle configuration** must be frozen, fully known and built as designed and its operational behaviour

Table 3.5.2: *Typical qualification tests to identify potential causes of failure or risk.*

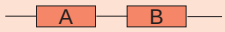
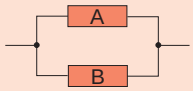
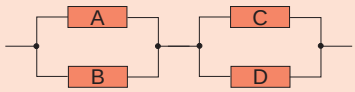
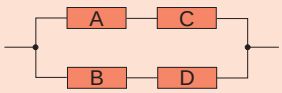
Potential failure mechanism	Primary qualification tests to identify failure mechanism									
	Functional	Vibration or acoustic	Shock	Thermal cycle	Thermal vacuum	Acceleration	Leakage	Proof and burst pressure	EMC	Life
Mounting broken/loose	x	x	x			x			x	
Broken part		x	x	x	x					
Shorted part	x	x			x				x	
Defective part	x	x		x	x				x	
Defective board	x	x		x	x				x	
Broken/shorted/pinched wires	x	x		x	x				x	
Defective/broken solder	x	x		x	x				x	
Contamination		x	x	x	x					
Leaky gaskets/seals/RF					x		x		x	
Incorrect wiring/routing design	x	x							x	
Relay/switch chatter		x	x						x	
Adjacent circuit board contact		x	x						x	
Premature wearcut		x								x
Electromagnetic interference									x	
Insufficient design margin	x	x	x			x		x	x	
Corona discharge/arcing					x					x
Inadequate tiedown of tubing/wiring		x				x			x	
Inadequate thermal design				x	x					
Brittle material failure			x							
Inadequate fatigue life		x		x						x

evaluated. On system level, performance and overall function must be validated, for instance via a stage hot firing test, and all open risks known and evaluated. All subsystems must have been accepted with respect to mechanical and functional usability. Materials, manufacturing and production processes have to be accepted. **Failure tree analyses** (failure mode, effects and criticality analysis, FMECA) must have been completed in order to understand the cause and the effect of potential failure chains from the component level up to the system level. Figure 3.5.2 shows the calculation principles behind analyzing failures occurring in series or parallel to others. A typical predicted theoretical reliability should be higher than 95% prior to the maiden flight.

Clearance for the maiden flight lift-off can be given when the required reliability has been demonstrated by the respective tests and statistical assessments and the operations for launch were finally verified.

The **failure of a launcher** might result not only in loss of the mission, but in a worst-case scenario, in the destruction of the launch facility or injuries to people. This explains the huge effort which is made on proving the correct and faultless operability of the overall system and the correct performance of all safety checks and measures before the first launch.

Figure 3.5.2: *Calculation of simple combinations and chains determining probability of failure.*

Type branch	Block diagram representation	System reliability #
Series		$R_S = R_A + R_B$
Parallel		$R_S = 1–(1–R_A)\cdot(1–R_B)$
Series-parallel		$R_S = [1–(1–R_A)\cdot(1–R_B)] + [1– (1–R_C)\cdot(1–R_D)]$
Parallel-series		$R_S = 1– [1– (R_A + R_B)] + [1– (R_C + R_D)]$

3.5.2 Categories of Qualification

The whole qualification process of a launch system not only comprises the launcher itself; it also comprises the **ground segment**, that is all the required facilities and operations for launch preparation and the launch campaign (see Section 3.4). For proving operational and mission readiness of the ground segment, in principle the same approach as for space flight equipment applies. However, other requirements and test methods are used to demonstrate the functioning of all installations and their coherency within the complete system; the communications network, all nominal and backup operational procedures, and last but not least the decision procedure's operational suitability. The qualification measures related to the ground segment are called verification of operational readiness.

Concerning the launch vehicle itself, the starting point for any qualification process is the **verification matrix**, which is subdivided into a number of requirement categories for better overview. These categories may vary depending on the specific project requirements; however, some typical categories for launch systems are:

- Mechanical system
- Thermal system
- Electromagnetic compatibility
- Mission
- Configuration
- Guidance, navigation and control
- Propulsion
- Power
- Communication
- Data management
- Flight operations
- Ground operations
- Support
- Human factors.

3.5.3 Mechanical Qualification

As an example, mechanical qualification is considered here in terms of mechanical tests to prove the structural readiness of the launcher. This also seems to be of particular interest, since due to the large size of a launch vehicle special requirements have to be defined to plan and to perform the qualification steps. Accordingly, qualification is often performed on complete launch vehicle stages.

The objectives of the mechanical qualification are:

1. To demonstrate the ability of the equipment to withstand the expected **static and dynamic loads** with a sufficient safety margin.
2. To validate that the **dynamic behavior** of the structure has been correctly characterized by analytical models, which is of particular importance for the flight loads analysis.
3. To demonstrate that the functional **structural components** comply with the required static and dynamic performance requirements with sufficient margins of safety.

4. To demonstrate that the operation of the **launch vehicle** is assured under the dynamic loads.

The relevant mechanical test methods (vibration test, static test, acoustic test, pyroshock screening and modal test) are explained in more detail in Section 8.3. In this context, however, the application of these test approaches to launch vehicles is discussed, addressing both mechanical strength and operational readiness.

Due to the size of launch vehicles, the respective qualification measures are practically restricted to **the main constituents of a launcher**, usually the launcher stages or major subsystems. Accordingly, within the scope of the complete system qualification approach, it has to be ensured that all subassemblies are separately qualified, in particular when their individual qualification is not possible on a higher integration level. For example: the qualification of a propellant line is mainly performed on the line itself, even if this line is later integrated into the stage qualification model.

To enable the testing of the launch vehicle main constituents or stages, it is mandatory that their **interfaces** and operational conditions are well defined and considered prior to testing. For example, since the propellant constitutes most of the launch vehicle's mass, the different **tank fill rates** and how they change during the mission have to be considered during testing with respect to the static and dynamic behavior. This may mean that several configurations, namely tank fill rates, have to be tested. Another example is that **stage separation** changes the boundary conditions from being restrained and clamped to free moving, and together with the changes in the propulsion during the staging sequence, the loads and fluxes alter significantly.

The size of the structures, the complex nature of the boundary conditions and the permanent changes in mass and enviromental conditions during flight require that mechanical tests be comprehensively planned, that representative **structural models** be provided at an early development step, and that considerable effort be undertaken to prepare or modify test facilities, in particular cases even to build new test facilities and to prepare dedicated jigs and loading devices. Consequently, the preparation and test evaluation phases are correspondingly long.

Taking the example of Ariane 5 launch vehicle development, some of the mechanical tests and their specific requirements and objectives are listed below.

Static Tests: Measurement of the fluxes, in particular close to the introduction points of the booster loads into the thin-walled structures; experimental verification of the stiffness matrices for the main and upper stage composites with respect to the fundamental compression and bending load cases; demonstration of the proof pressure of the stage propellant tanks; verification of the ultimate load and the damping characteristics of the elastic link elements between the boosters and the main stage under compression loads combined with thermal and dynamic stress.

Vibration Tests: Dynamic qualification of the main engine, since the engine's operation loads are supplemented by the vibration loads occurring at the lower section of the main stage; vibration test on the upper stage to verify the dynamic strength and to investigate the vibration and damping characteristics of the upper stage tanks; experimental verification of the effectiveness of the damper system of the upper stage's main tank. Figure 3.5.3 shows the ESC-A upper stage during its vibration test.

Pyroshock Tests: Characterization of the propagation of shock waves and body noise within the upper stage caused by stage separation and separation of the acceleration rockets. This is also determined for the separation of the payload and for fairing jettisoning.

Acoustic Tests: Determination of the acoustic noise attenuation of the nose fairing; characterization of measures to reduce the acoustic load inside the payload volume; acoustic noise test on the upper stage to verify the acoustically induced vibrations at the location of the electronic flight control systems of the launcher.

Modal Tests: Determination of the vibration characteristics of all main launcher components (booster, main stage, upper stage including the payload section) for a number of different flight conditions, tank fill rates and boundary conditions (restrained–free or free–free); dynamic characterization of intertank structure damping devices within the upper stage; experimental verification of the damping characteristics of the elastic link between booster and main stage under full compression load and dynamic stress;



Figure 3.5.3: *Vibration test of the Ariane ESC-A upper stage (Source: IABG).*

investigation of the dynamic responses at the location of the inertial platforms due to modal vibration.

Functional Tests under Dynamic Stress: Dynamic behavior of the upper stage during the transient processes during engine ignition and shutdown; operation of the tank pressurization system and pressure control under acoustic excitation, POGO excitation of the

- feeding system due to engine combustion.
- propellant.

These examples may be sufficient to show the variety of mechanical investigations and qualification measures. Finally, it should be noted that some parameters *cannot* or *can only to a certain extent* be realized in a laboratory environment. For instance, the use of real propellant is often prohibited because of the cryogenic temperatures, explosion hazards and toxicity. Therefore, the differences in mass density of the replacement fluids has to be considered. Application of the real flight pressure is also impossible due to the inherent risk of a tank rupturing during the test. These differences, as well as mechanical boundary conditions which can only be realized in an approximated sense, must be taken into account in the mathematical models used for the test prediction. Accordingly, during the first flights additional measurements are performed to provide further evidence for the qualification status.

3.5.4 Functional Qualification

Functional qualification comprises thermal, thermal–functional and propulsion system verification. **Thermal engineering** determines the thermal fluxes into, and temperatures of, the structures, components and propellant. Moreover, the thermal design of the launcher is specified. Specific challenges are thermal fluxes due to aerodynamic heating, aerothermal interaction with the engine exhaust, and determination of the propellant temperatures inside the tanks (in particular when considering interactions with fluid dynamic effects). The role of thermal engineering within the verification process is to determine a representative thermal environment for structural and thermal–functional tests, to determine test conditions and sensor installations, to analyze the thermal behavior during test, to investigate thermal anomalies which occur during flight, to assess and correlate ground and flight test data with analytical thermal models, and to review the efficiency of the thermal design.

Thermal–functional verification deals with tests for filling and draining propellant fluids and, applicable to cryogenic propellants, determining the temperatures and pressures during these operations. Further activities are checking the pressurization procedures for the propellant tanks, filling and venting the tank pressurization systems, determining pressure loss in lines and investigating dynamic flow effects, determining suitable mass flow rates for the ventilation of cavities inside the stage, investigating gas flow rates for the local conditioning of equipment, detecting and analyzing leakage, determining expansion flow rates for the evacuation of lines, and reviewing the efficiency of these operations and the functional design. Finally, the analytical models are correlated with the experimental results and updated, when required.

The **qualification of the propulsion systems** considers both the transient and stationary operational phases of the engines. In doing so, each single component, like pumps, igniter and thrust chamber, as well as the complete engine are verified. The test involving the whole engine is called the **firing test**. When involving the entire propulsion system and the stage, the test is called **stage hot firing test**. The preparation of hot firing tests requires the intense cooperation of all engineering disciplines involved, since, for instance, the gas flow in the engine may excite significant structural vibrations at the engine nozzle or the feed lines. The running engine causes high thermal stress (thermal radiation, hot exhaust gas flow) for the engine, the test stand and the environment as well. It is quite complex to operate on the upper stage under real vacuum conditions on the groŭnd. Therefore, in so-called **development tests**, the behavior of engine components and partial engine working cycles are verified, applying water or gas flow tests. During **engine qualification** all relevant operational modes and functions (operational box) are tested, in particular the transient phases like ignition and run-up, as well as the shutdown phase. Instable conditions during ignition or later during nominal combustion may quickly lead to engine loss, since the combustion of the propellant is a highly energetic process. Irregular expulsion of gas may lead to hot spots which may melt the material of the nozzle throat. Hot firing tests on main stage engines or booster engines have to consider that high vibration and acoustic loads may occur and easily lead to fatigue failures on the nozzle or other engine parts. If the engine uses regenerative cooling, the correct function of this system must of course be shown before any hot firing takes place. Finally, for cryogenic engines specific operations like the cooling phase of the lines and components down to the propellant temperature (chilling down) have to be verified. A particular testing problem with main stage engines arises from the fact that during the ascent phase they are operated against decreasing atmospheric pressure. Tests must accordingly consider this fact. A similar problem exists with upper stage engines which are ignited and operated in vacuum. Consequently, the final qualification of an engine usually ends with the qualification flight.

Further Activities after Qualification

Once the qualification flight is completed, the launch vehicle enters the consolidation phase. Results and insights from the qualification flight might lead to necessary changes in hardware and software. As a further consequence, the flight measurement equipment is reduced step by step to the essential minimum for the next flights. For the beginning of serial production, the configuration of the launcher must be frozen in order to achieve a generic qualification.

Then the **production process** is stabilized and, whenever required, optimized. From time to time the manufacturer of a certain part may change during the production phase. This requires the provision of specific acceptance tests to ensure **acceptable** quality according to the specification after the change of the production source. In this phase special attention is given to manufacturing alterations, **obsolescenses**, and flight anomalies. It has to be proven in a continuous process that even in the case of alterations the complete system still remains qualified within its nominal operations with respect to the specifications.

If the launcher system has to be adapted to **specific mission profiles** which are not fully covered by the generic qualification, a delta qualification of subsystems or even the whole system might be required.

Discovery

4 Subsystems of Spacecraft

Wilfried Ley and Fritz Merkle

The tasks that a spacecraft has to perform – or, in other words, the mission – determine its configuration. Based on a thorough analysis of the definition of the space mission, the requirements for the overall system, consisting of the space, transfer and ground segment, have to be derived.

The overall system requirements represent the fundamentals for the configuration of the above-mentioned system segments, which are structured in system elements (see also Section 1.2). The space segment consists of the system elements "spacecraft" and "payload." In this chapter the major and typically required subsystems of the system element "spacecraft" are described.

The spacecraft carries the payload and has to be adapted to the payload's requirements (see Section 1.2). In the case of a satellite this is often called the satellite bus or platform. This results in a number of typical configurations for the spacecraft. Examples are shown in Figure 4.1 for a spin-stabilized satellite, a three-axis-stabilized satellite (today the most common configuration) and a freely configurable spacecraft adaptable for special tasks.

Independent of its configuration, a spacecraft is composed of a combination of typical subsystems, which will be elaborated in this chapter.

The structure and the associated mechanisms influence significantly the configuration of the spacecraft and have to be frozen at an early phase of a project since a major change is usually not possible at a later phase. In this case, the configuration of the spacecraft would have to be reworked almost from the beginning. The mechanisms discussed in the first section focus on those which are linked closely to the structural elements of the overall system. In other words, the structure is the body of the spacecraft into which the other subsystems are integrated and which carries the static and dynamic loads during the launch of the spacecraft.

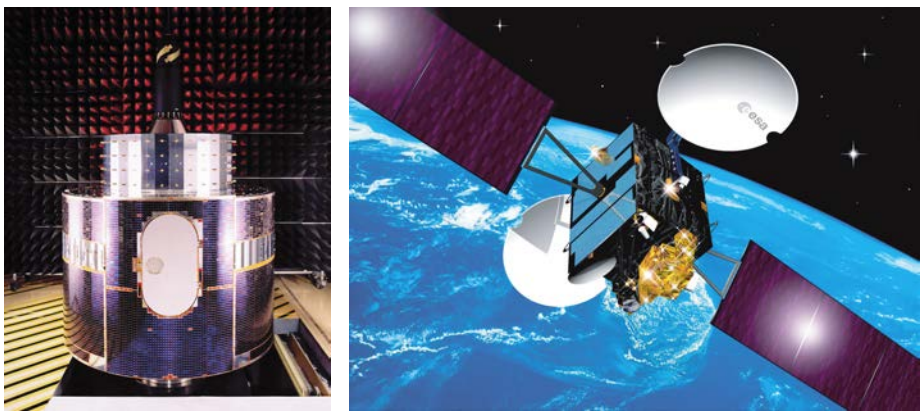


Figure 4.1: *Spacecraft configurations. Example of a spin-stabilized satellite (Meteosat, left), a three-axis-stabilized satellite (Artemis, right) and a freely configurable spacecraft (the Space Shuttle, facing page) (Source: ESA, NASA).*

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9

The power supply required to operate the payload and the platform with its subsystems and the communication and data transmission services is discussed in the next section. Elements of the power supply are power generation, power conditioning, power storage and power distribution inside the spacecraft. Apart from typical solar power generation, nuclear and chemical power generators are introduced, as well as the most familiar battery types for power storage. The design of the power supply and the availability of this supply throughout the entire mission are of essential importance for a space mission.

As soon as a spacecraft has started its journey it is exposed on the one hand to the cold temperatures of space and on the other hand to the direct unfiltered radiation of the Sun. This results in high temperature fluctuations with the consequence of steep temperature gradients, which have to be taken into consideration. Most of the materials and components used in the construction are only suited to more moderate and Earth-like temperature ranges and often withstand only very limited gradients.

In addition, all active subsystems in the spacecraft generate heat, which has to be included in the considerations as well. Therefore, an environment must be generated within the spacecraft which allows all subsystems and components to be operated within the specified and acceptable temperature tolerances. To achieve this, the temperatures have to be monitored and controlled by passive or active methods inside the spacecraft. This is explained and shown by realistic examples.

The first spacecraft were carried directly into their target orbit by a rocket. They lacked their own propulsion system. Today, most spacecraft are equipped with a propulsion system to maintain or change their orbit. This is usually achieved by propulsion systems based on thermodynamic or electrodynamic processes. Depending on the mission and propulsion type, the propellant quantities have to be adjusted. For example, to change from a transfer orbit to a geostationary orbit the required propellant quantity is nearly half the mass of the satellite itself, in the case of chemical propulsion. In most cases, the propulsion systems serve to maintain and correct the orbit or to initiate planned orbit maneuvers. The use of electrical propulsion systems reduces the required propellant quantities significantly. The on-board propellant quantities and their economical use are essential aspects of the duration of a mission. At the end of a mission, a certain quantity of propellant must be conserved so that the spacecraft can perform controlled deorbiting or vacate its orbital position.

Besides the precise positioning of a spacecraft in an orbit, often the attitude of the spacecraft is of importance for the success of a mission. Early spacecraft were stabilized by rotation around their own axis. Today, most satellites have to be operated in precisely defined and maintained orientations, or, as in the case of Earth observation satellites, they have to point precisely at a selected target on the ground. To achieve this, an attitude control system has to determine and adjust the orientation. In the section describing attitude control, it will be shown how a selected attitude and motion can be achieved, starting from an initial attitude and motion. Sensors like gyros, accelerometers, star sensors, GPS, etc., and actuators like torquers and electromagnetic coils are applied for the purpose.

Together with the payload, the various subsystems form a network which is controlled by one or more on-board processors. Often payload and platform tasks are distributed between two or more subsystems which communicate with each other. The processors are interconnected among each other and with all measurement devices, sensors, actuators and other modules of the spacecraft. The control system is responsible for achieving maximum autonomy for the spacecraft. This requires not only high computing power, but often also very high software complexity. In addition, this subsystem has to be highly reliable, fault tolerant and if possible self-healing – and all within a small volume with low mass – and still be operational under the extreme environmental conditions of space.

A spacecraft can only perform its mission if it is linked to a ground station for control and data reception. Depending on mission requirements, a space system needs a communication system with a reliable and ground-station-compatible transmission capability, to guarantee a stable link to the Earth.

The various subsystems of a satellite cannot be regarded independently of each other. They are linked and interconnected and influence each other.

Therefore, the subsystem definition is followed by an iterative process to achieve optimization of the overall system by using optimized subsystems (see Section 1.2 and Chapter 8). Thus the success of a space mission depends on the complex interoperation of all subsystems, which is described in more detail in the following sections.

4.1 Structure and Mechanisms

Joachim Block and Johannes Kreuser (4.1.5.6, 4.1.5.7)

In organizational charts and listings displaying the subsystems of a spacecraft the **structure** appears very often in first place, usually followed by power supply and thermal control. This first place has less to do with any "ranking" than with the sequence in which the subsystems become design drivers. Like the hull of a ship in the shipbuilding process between keel laying and launching, the structure of a spacecraft determines the overall configuration already at an early phase of the development process, and also needs to be qualified first. Consequently, the first qualification model of a satellite is usually the **structural–thermal model** (STM). If it does not satisfy the requirements, then possibly the whole configuration and all the mechanical interfaces will need to be redesigned from scratch. This later becomes almost impossible unless years of delay are taken into account. On the other hand, a redesign within the electronic subsystems, for instance, is in most cases still possible at a later stage of the design process without endangering the course of the project as such.

The discussion of **mechanisms** in this chapter focuses on those which are directly related to the structural mechanics of the overall spacecraft and which therefore need to be considered in conjunction with the structure. This comprises all mechanisms which serve for the locking, releasing, deployment and active mobility of essential spacecraft components (e.g., antennas, solar cell arrays, deployment booms) relative to the primary structure.

4.1.1 The Primary Structure of the Spacecraft

The load-carrying cell of a spacecraft, which determines its global **strength** and **stiffness**, and upon which all other units are fastened, is defined as the **primary structure**. From the stiffness of the primary structure and the mass distribution of the various system and payload units mounted on it, the **resonance behavior**, that is the **eigenfrequencies** under dynamic load, can be calculated. Normally an approximate calculation of the eigenfrequencies is already required in the second phase of a project in order to check compatibility with the specifications of the envisaged launcher (rocket, shuttle).

4.1.1.1 Design Drivers and Dimensioning Load Cases

As illustrated in Figure 4.1.1, the first, still rough concepts for the basic structural configuration of a spacecraft emerge at the beginning of the project in the first phase, Phase A, based on the objective of the mission. For instance, a free-flying satellite with a mainly electronic payload looks principally different from a space telescope, which is designed around a tube, or from a planetary lander with landing gear or airbags. The requirements of the mission-determining main payloads act as **design drivers**, in the beginning mainly as geometry drivers, and must be made compatible with the constraints of the envisaged launcher. In cases where the volume under the fairing must be shared with other satellites, even these constraints must be considered from the beginning.

This initial draft specification of the spacecraft structure is then, in the course of the subsequent project phase, Phase B, subjected to thorough numerical analyses by means of a first **finite element model**. These analyses prove the basic feasibility of the structural concept and exclude potential "showstoppers" as long as the construction process is still flexible

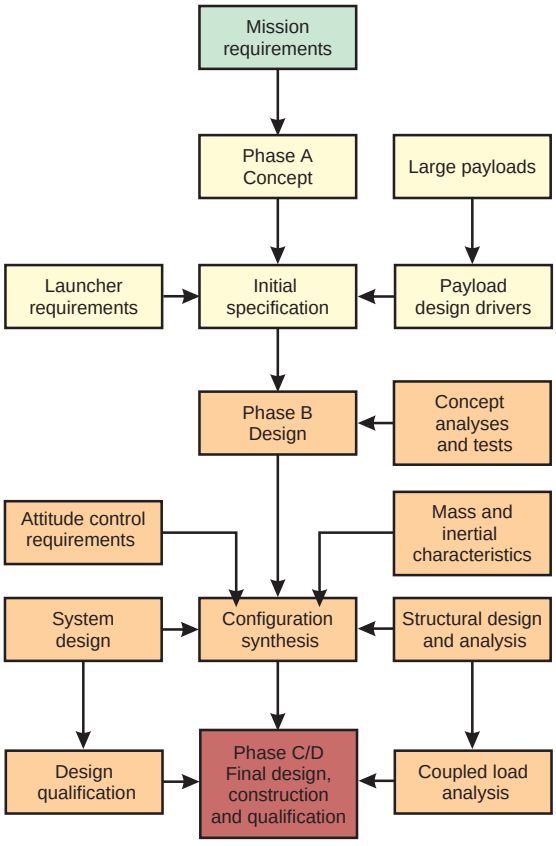


Figure 4.1.1: *Typical structural design process (Source: [4.1.1]).*

enough to allow for major modifications. Also, testing of **breadboard models**, that is of hardware which does not yet count as official qualification hardware and is not yet subject to any configuration control, can decisively strengthen confidence in the selected structural concept and help to avoid "bad surprises" during the STM tests, in particular if important aspects of the structural design cannot be realistically modeled in advance.

Afterwards, the effects of the installed payload and system units on the **mass distribution** and on the **moments of inertia** are iteratively considered with increasing accuracy, as are the attitude control requirements, so that finally the overall configuration at the system level can be reliably modeled and numerically analyzed.

In the project phase, project C/D, this process finally culminates in the consolidation of the final design and its experimental qualification, which is usually performed with the STM. Subsequently the other hardware models up to the flight model (FM) are built.

The **dimensioning load cases** which must be considered for any mission include in any case the **dynamic loads** during the launch phase. The **sinusoidal**, **random** and **acoustic load** spectra generated by the rocket engines are, in conjunction with **quasi-static** accelerations, the main challenge for the structure as a whole, whereas **shock loads**, as occur when pyrobolts are fired (e.g., when the stages are separated), are mainly critical only for sensitive mechanisms.

Since it is not possible in the beginning of the design process to predict in detail the effect of the various dynamic loads on the structure, as well as on the payload and system units mounted on it, a **quasi-static design load** is predefined (also called the **design limit load**), often in the form of a logarithmic plot. It is calculated so that the anticipated dynamic loads are respected with the utmost probability. Figure 4.1.2 shows as an example the requirements for the BepiColombo science mission [4.1.2].

Associated with the mass of each unit on the spacecraft, an acceleration is defined which must be "sustained" and which serves as target parameter for the dimensioning process, usually combined with

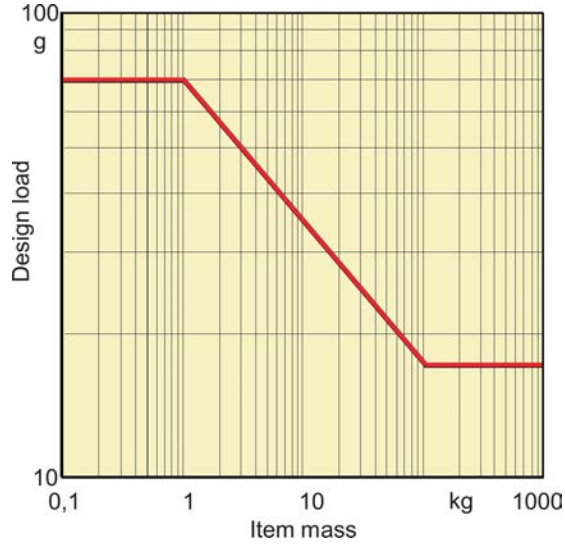


Figure 4.1.2: *Requirements for a quasi-static design load.*

Table 4.1.1: *Selected material properties (averaged values from different manufacturers).*

Structural material	Strength under tension	Stiffness	Density	Thermal expansion
Unidirectional laminate with epoxy matrix:	(in each case with 60% fiber volume content)			Low
• With high-tenacity (HT) carbon fibers divided into high (HS) and very high (VHS) strengths	HS: ≈ 2100 VHS: > 2100	120–150	1500	< 0 in fiber direction, > 0 normal to fiber direction In fabric approx. 0
• With high-modulus (HM) and ultrahigh modulus (UHM) carbon fibers	1400–2300	HM: > 310 UHM: > 395	1600	
• With aramid fibers (Kevlar/Twaron)	1400–2000	≈ 75	1400	
• With glass fibers (S-glass)	≈ 2000	≈ 50	≈ 1800	$5 \cdot 10^{-6}$/K
Aluminum (HT alloys)	310–450	≈ 70	2750	$23 \cdot 10^{-6}$/K
Titanium	700	110	4450	$9 \cdot 10^{-6}$/K
Beryllium	320	300	1860	$11 \cdot 10^{-6}$/K
AlBeMet (with 62% beryllium)	262	193	2100	$14 \cdot 10^{-6}$/K
Carbon/carbon	200–260	≈ 90	≈ 1850	Very small

additional **safety factors** (see Section 4.1.3.3). For the primary structure, which must be able to carry all integrated units, it is of course the total mass of the spacecraft which determines the quasi-static design load.

During the free cruise mode in space the comparatively compact primary structure is no longer subject to any mechanically critical loads. Also the mass forces induced by trajectory correction maneuvers are in almost all cases negligible. However, the situation is different for deployed large secondary structures, such as unfolded booms or panels with solar cell arrays or large antennas (see Section 4.1.2). The handling loads under Earth's gravity which occur during integration and all ground transport activities ought to remain uncritical under all circumstances. This has to be assured by appropriately designed **mechanical ground support equipment** (MGSE).

For extraterrestrial missions completely different load cases may govern the dimensioning process. In the case of an airbag-supported landing of a space probe on Mars, for instance, the transient mechanical load upon impact may be more critical by far than all launch loads. For exploration missions into regions of the Solar System which are either close to or far from the Sun, the thermal requirements may become much more decisive for the primary structure than the mechanical ones, particularly if different structural parts need to be effectively thermally isolated from each other. Also, thermoelastic stresses then need to be considered.

4.1.1.2 Selection of Structural Materials

In order to satisfy the mechanical and thermal requirements for the spacecraft structure with the lowest mass possible, it is decisive to appropriately select the structural materials. The following material classes are relevant for this purpose [4.1.3]

Metallic Alloys

As in aircraft construction, the classic **aluminum alloys** still hold the leading position among the metallic structural materials in the field of spacecraft construction, followed by **titanium-based alloys** [4.1.4], [4.1.5]. In parallel, novel materials are gaining importance for special applications, for instance **beryllium alloys**, which are extremely light in weight and have excellent mechanical properties, but on the other hand are very problematic to process and manufacture, in particular because of their extreme toxicity.

Apart from small parts (e.g., titanium screws) the semifinished products used for the manufacture of metallic spacecraft structures are mainly sheets, plates and profiles. Manufacturing from solid metal, for instance the monolithic manufacturing of electronic boxes, is usually performed by countersinking with a computer-controlled fraise (see Section 4.1.1.3).

Fiber Composites

The development and steady improvement of **carbon fibers** [4.1.6] over the last few decades has paved the way for the leading role of **carbon fiber-reinforced plastic** (CFRP) laminates in modern aerospace engineering. Laminates composed of high-modulus or high-tenacity carbon fibers embedded in a thermosetting matrix (mostly epoxy resin) are today the backbone of the structural technology for space applications [4.1.7], [4.1.8], [4.1.9], [4.1.10]. Laminates with thermoplastic matrices, on the other hand, are of minor importance in space due to their lower heat resistance.

Flat laminates are frequently manufactured from pre-impregnated fiber layers (**prepregs**), while for structures with large curvatures and more complex forms it is better to first position the still dry carbon fiber layers (fabric etc.) in a mold and then impregnate them with the still-liquid matrix resin. For the manufacturing of structures with approximately rotational symmetry, such as cylinders or tanks, the dry fiber roving is pulled through an immersion basin with liquid epoxy before it is wrapped around the rotating preform (**filament winding technique**). In all cases the hardening and curing are performed in an autoclave.

Occasionally **glass fibers** or **aramid fibers** are also used (the latter consisting of aromatic polyamides and sold under such trademarks as **Kevlar**, **Twaron** or **Nomex**), in particular when the good electrical and thermal conductivity of carbon fibers is undesirable.

One of the most common combinations of fiber composites with metals in the same structure are sandwich plates with CFRP face sheets and an aluminum honeycomb core (see Section 4.1.1.3). For many applications this combination represents the optimal solution.

Fiber–Metal Composites

This material class comprises **hybrid materials** such as **fiber–metal laminates** (FMLs), which consist of thin metallic sheets that are mixed layer by layer with carbon fiber, glass fiber or aramid fiber layers (GLARE, ARALL, etc.), as well as **metal matrix composites** (MMCs), which combine the mechanical advantages of fiber composites with a much higher temperature resistance, at the price of higher weight. They already have a bridging function to composites that are decidedly high-temperature materials.

High-Temperature Materials

For extremely high temperatures as occur during hypersonic flight through the atmosphere, for instance during the reentry of envisioned reusable space vehicles, materials made of **ceramic matrix composites** (CMCs), **carbon/carbon compounds** (C/C) and **silicon carbide compounds** (C/SiC) are being developed.

Table 4.1.1 gives a global overview of the most important properties and parameters of all these materials. A more detailed characterization of the hundreds of individual material specifications is not possible

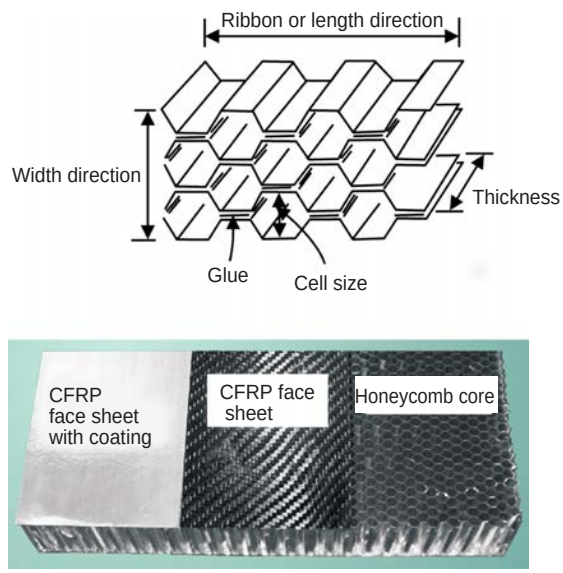


Figure 4.1.3: *Composition of a sandwich with a honeycomb core (Source: DLR).*

within the frame of this handbook; the interested reader could consult the references in the "Structural Materials Handbook" [4.1.3] edited by ESA and ECSS, or pertinent material handbooks [4.1.4], [4.1.5], [4.1.6], [4.1.7], [4.1.8], [4.1.9], [4.1.10].

4.1.1.3 Sandwich, Differential and Monolithic Construction

Sandwich Construction

Today, sandwich construction is the most widely used approach for manufacturing spacecraft structures. Sandwich plates consist of two parallel **face sheets**, preferably aluminum alloys or carbon fiber laminates, which are glued onto both sides of an intermediate **sandwich core**. In most cases the latter is a **honeycomb core** with hexagonal cells, but corrugated metal sheets, sheets with a zigzag profile, or hard foam cores are used as well.

The preferred core material is aluminum, while aramid fiber paper (**Nomex**), which is quite common in aircraft construction, is not often used in space applications because of the environmental conditions.

Figure 4.1.3 shows in principle the composition of a sandwich plate with a honeycomb core with hexagonal cells. Its mechanical properties, in particular its **strength** and **stiffness** under tension, compression, shear and bending loads, depend decisively on the face sheet thickness, the core thickness, the cell size and the thickness of the honeycomb cell walls – apart from the pure material properties as such. Here it must be noted that, due to the manufacturing of the honeycomb core from individual thin ribbons which are folded and then glued together at the points depicted in Figure 4.1.3, the mechanical properties in the length direction differ from those in the width direction.

Since adhesion between the honeycomb core and the sandwich face sheets is achieved by means of an adhesive film which creeps into the fillets between the inner side of the face sheet and the cell walls perpendicular to it and hardens there, the cells would be hermetically closed if **venting** were neglected. By perforating the ribbon material in the course of honeycomb core manufacturing it is assured that after launch of the spacecraft the air enclosed in the cells can quickly evacuate through small holes or slits.

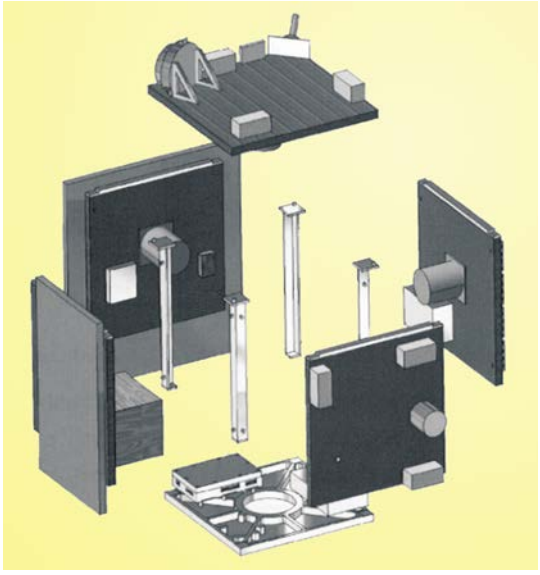


Figure 4.1.4: *Small satellite structure composed of sandwich plates (Source: DLR).*

The large number of sandwich parameters (particularly if the face sheet is not isotropic but a multidirectional fiber laminate) makes it possible to "tailor" a variety of sandwich properties, but makes their analytical calculation rather complex. It would go far beyond the scope of this handbook to derive and discuss all the corresponding formulas from some general basic equations down to a level practically usable for **sandwich analysis**. Therefore, reference is made to some excellent books on the subject, such as those by Plantema [4.1.11], Allen [4.1.12] and Altenbach *et al.* [4.1.13]. For the dimensioning of fiber composite face sheets see Michaeli *et al.* [4.1.14].

Spacecraft structures can be composed of sandwich plates in different basic geometrical forms. The simplest case is illustrated in Figure 4.1.4 [4.1.15] and is representative of many small satellites. Their basic form is a cube or cuboid where five of the six sides are lightweight sandwich plates and the sixth is a more massive bottom plate. This configuration can be easily folded together and is therefore easy to integrate as well. All payload and system units are fastened onto the inner sides of the side walls; no internal structure is needed.

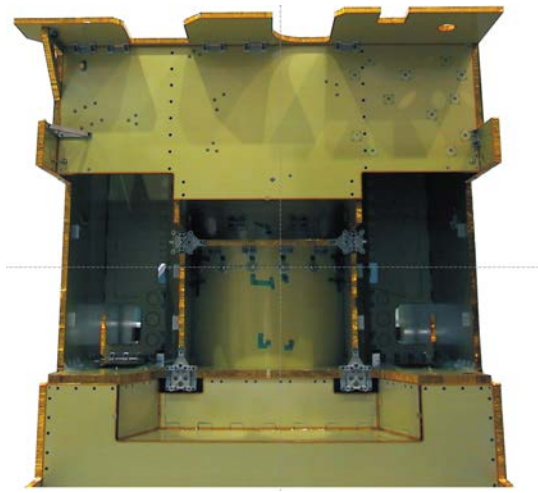


Figure 4.1.5: *Structural configuration of the Rosetta orbiter (Source: ESA).*

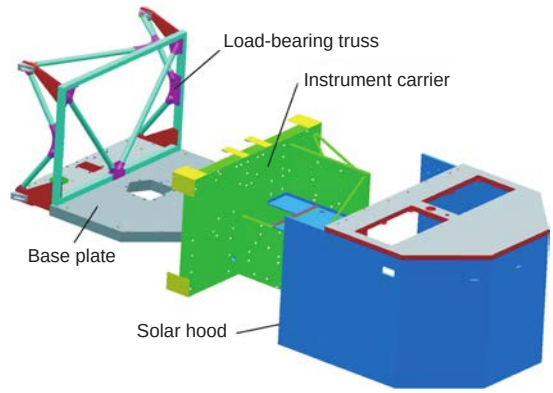


Figure 4.1.6: *Structural configuration of the Rosetta lander* Philae *(Source: DLR).*

4

Large satellites, however, need a central structure. An often favored solution is a combination of a central tube with a circular cross-section, which can be advantageously manufactured using a **filament winding technique** and a cuboid outer sandwich structure. With an appropriate liner inside, the central tube can be ideally used as a fuel tank. The surrounding sandwich plates form several individual compartments for accommodation of the payloads. Figure 4.1.5 shows a half-finished structural model of the orbiter for ESA's comet mission Rosetta. Since the middle sandwich plate on the front side is still missing, one can recognize the central tube and the "hard points" for later fixation of the Rosetta lander. These are located at the intersection of two (finally three) sandwich plates oriented perpendicular to each other.

Different from this classical configuration is the primary structure of the Rosetta lander, as shown in Figure 4.1.6. It is a good example of a sandwich structure, the design of which was not primarily driven by mechanical requirements alone, but more by thermal design drivers. After separation from the orbiter, the lander is destined to land on the surface of an active comet, in an extremely cold environment. Consequently, all temperature-sensitive payload and system units are fastened onto an instrument carrier, which is itself composed of several sandwich plates and completely wrapped with multilayer insulation (MLI) blankets. The only rigid connections from this instrument carrier to the external structure, that is to the base plate and the load-bearing support truss, are thin-walled Kevlar elements which penetrate through slits in the MLI and are aligned along the major load paths through which the mass forces are transferred to the four fixation points (hard points) on the orbiter. The whole configuration is covered by the solar hood that carries the solar arrays, some cameras and absorbers, and is an integral part of the load-bearing primary structure in spite of its extremely lightweight design.

In some satellites the sandwich plates are arranged one level over the other, connected by a truss of stiff rods. In this case the truss has to ensure not only high stiffness along the longitudinal axis of the satellite, but also high torsional stiffness. An example is shown in Figure 4.1.7.

Differential Construction

Differential construction is the most traditional form for lightweight structures. In the early years of the space age most satellites were built "differentially": that is, by combining simple preformed components made of conventional materials, for instance shells and profiles made from aluminum sheets which were subsequently connected by rivets or screws, similar to aircraft construction of that time.

Although these preformed components are very easy and cost effective to manufacture, even in large



Figure 4.1.7: *Structure of the meteorological satellite MSG-1 (Source: ESA).*

quantities, this aspect is less relevant for spacecraft construction. On the other hand, their integration is very time consuming and expensive, in particular if hundreds or thousands of rivets (or screws, or adhesive connections) are required to complete the structure as a whole (Figure 4.1.8).

The advantage of being relatively resistant to **crack propagation** (because cracks end at the edges of the individual components or at the bearings of the screw or rivet holes and do not propagate through the entire component) is offset by the relatively high weight. Therefore it is not surprising that during the last few decades, in view of the growing importance of modern fiber composites on the one hand, and the development of modern computer numerical control (CNC) manufacturing techniques on the other hand, the traditional methods of differential construction have taken a back seat.

Monolithic Construction

The philosophy pursued in monolithic (or "integral") construction is just the opposite of the approach used

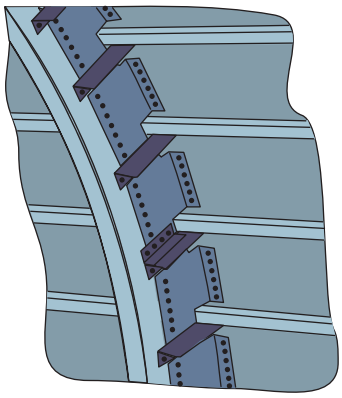


Figure 4.1.8: *Skin and stringer in a differential construction.*

in differential construction: ideally the whole structure consists of a single monolithic piece, or is at least composed of only a few large parts.

If monolithic **metallic structures** are considered, the only really feasible production technique is manufacturing from the solid by means of **CNC tools**. Such monolithic parts are much cheaper and easier to produce today than in the early days of the space age, and also the achievable precision is much higher. The theoretical alternative for producing monolithic metallic parts by casting is practically irrelevant for space structure technology, apart from rare exceptional cases.

Certainly it is true that for large dimensions, structures composed of sandwich panels are clearly lighter in weight than corresponding "integral" monolithic structures because milling and grinding from the solid material cannot be accomplished to achieve such extremely thin wall thicknesses as are possible with sandwiches. This is also true for so-called **isogrid structures** where panels are stiffened on their back side by a monolithically connected network of stringers (ribs), preferably with equilateral triangles in between (i.e., all angles are 60°). For smaller and more compact structures, however, the monolithic construction method is well established. Indeed many payload and electronic boxes are today manufactured from an aluminum solid by means of a CNC machine and subsequently surface treated, in particular chromated ("iridite").

In the case of fiber composite structures the monolithic construction method remains competitive

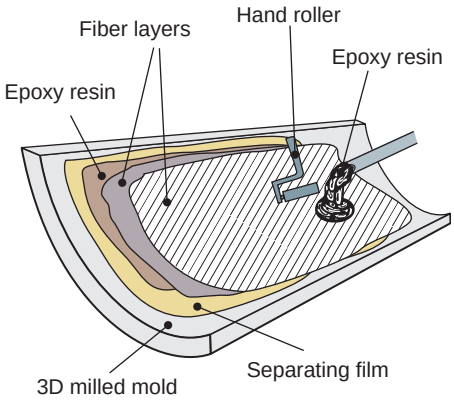


Figure 4.1.9: *Hand lamination of fiber composite structures.*

with sandwich construction even for larger dimensions. This is particularly true for curved shells, since sandwich panels with a one-dimensional curvature are difficult to manufacture, and even more difficult if the curvature is two dimensional (and the basic core material was initially flat).

On the other hand, there is no problem at all in precisely manufacturing a negative form curved in two dimensions using a CNC machine and to use it subsequently as a lamination mold. As long as the required wall thickness is no larger than several millimeters, skilled personnel can manufacture structural parts of high quality on such a mold even by simple hand lamination (Figure 4.1.9), because the homogeneous interpenetration of fiber material and liquid lamination resin (epoxy) can be directly observed. The readily hardened curved shell can later be stiffened with stringers that are glued on.

For structures with thicker walls, however, automatically controlled injection techniques are required, such as the different varieties of **resin transfer molding** (RTM) [4.1.16], [4.1.17], [4.1.18]. In this technique the flux behavior and the viscosity of the resin are actively controlled, and precise pressure levels are used to make sure that the fiber layers or fabrics are homogeneously wetted by the lamination resin even to large depths. **Hardening** in an autoclave is also decisive for the quality (Figure 4.1.10). In this field significant improvements can be expected from wider future use of microwave-heated autoclaves [4.1.19].

An example of a monolithically manufactured space probe structure is shown in Figure 4.1.11. It is



Figure 4.1.10: *CFRP panel hardened in an autoclave (Source: DLR).*

a CFRP shell designed for the instrumented lander capsules of a former Mars exploration project (NetLander, [4.1.20]). The shell is a high-modulus carbon fiber fabric laminate of only 3 mm thickness which survived impact tests with the whole 33 kg mass of the capsule up to 180*g* without damage.



Figure 4.1.11: *Monolithic shell of the NetLander probe (Source: DLR).*

4.1.1.4 Outgassing, Degradation and Surface Protection

A particular challenge for spacecraft structures is their endurance under the environmental conditions of space (for details, see Section 2.1). The main influences are:

- The omnipresent **high vacuum**
- **High-energy radiation** from the Sun and from the galactic background (ultraviolet, X-ray and gamma radiation, see Figure 2.1.2)
- **High-energy particles** (protons, neutrons, electrons, alpha particles) from the solar wind and in the **Van Allen belts** of the Earth
- **Atomic oxygen** (in low Earth orbits).

The high vacuum leads to a mass loss of all structural materials which contain volatiles. This does not concern metals, but all plastic materials including fiber composites. Since the corresponding external vapor pressure is zero, these materials lose their volatiles, primarily water molecules, by **outgassing**. This dehydration may not just lead to critical embrittlement of the material itself. Even more critical is the subsequent condensation of the volatiles on other parts of the spacecraft, for instance on camera lenses, mirrors and other sensitive sensor surfaces, which can blind their field of view.

Spaceborne structural materials are therefore subject to strict outgassing criteria which are summarized in a pertinent ECSS standard [4.1.21]. Compliance with this standard must be verified by outgassing tests (Figure 4.1.12).

After a preceding conditioning phase under normal ambient temperature and humidity, a sample of the material is heated for 24 hours at 125 °C in a vacuum so it can outgas. Immediately prior to and after the outgassing phase the sample is weighed so that the **total mass loss** (TML) can be determined. During a following second conditioning phase the sample is again saturated with water molecules (**water vapor regained** (WVR)) so that the **recovered mass loss** (RML) can be determined by the difference in weight. In addition to this procedure a sensitive collector plate in the test device is used to measure the amount of solid particles which are emanating from the sample during the outgassing phase (**collected volatile condensable material** (CVCM)). If not otherwise specified, the acceptance criteria for spaceborne materials are TML < 1% and CVCM < 0.1%. In the vicinity of optical instruments even RML < 0.1% and CVCM < 0.01% are often required [4.1.3].

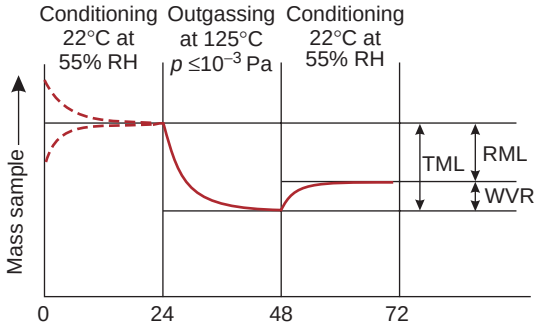


Figure 4.1.12: *Outgassing test according to ECSS-Q-70-02A.*

In view of the material **degradation** caused by ultraviolet radiation and high-energy particles which may gradually destroy molecular bonds in the material, it is necessary to estimate the lifetime dose to which the spacecraft will be exposed. Apart from the mission duration itself, this dose depends of course on the local environment; that is, on whether the spacecraft is operating in a low Earth orbit, beyond the Van Allen belts, in deep space far from the Sun, etc. Also, the possible coincidence of the mission duration with phases of strong solar activity within the 11-year cycle of the Sun may be significant.

Again fiber composite materials with their polymeric matrix are more vulnerable to the destructive influence of high-energy photons or particles [4.1.22] than are metallic alloys. Protection of their surfaces is therefore mandatory. Common metallized Kapton foils or MLI blankets which are primarily used as thermal insulation for satellites and space probes provide considerable protection against short-wave electromagnetic radiation, even though high-energy particles may still penetrate through these covers. Another option is to laminate thin protective metallic foils directly on carbon fiber laminate surfaces. Due to the different thermal expansion coefficients one might expect clefts or creases in the foil, but it is possible to avoid them (see the covered face sheet in Figure 4.1.3). A preceding chemical or electrochemical

treatment of the metallic foils (e.g., the so-called "iridite" procedure) prevents the surface being covered by a nonconductive oxide film.

A conductive and well-reflecting metallic surface not only is a very valuable protection of the structural material below against degradation, but moreover prevents local electric charge concentrations on the structure. The problem of **surface charging** is particularly important in a geostationary orbit where the plasma environment may lead to strong local charge differences between different parts of the spacecraft [4.1.23]. Voltages up to several kilovolts can be generated and lead to sudden destructive discharges if they are not prevented by proper "grounding" by conductive connections among the different surface areas.

Finally, for missions in a low Earth orbit it should be considered whether **atomic oxygen** might have a degrading effect (for all other trajectories this can certainly be neglected from the outset). Atomic oxygen is generated by the dissociation of oxygen molecules in the highest parts of the atmosphere by solar UV radiation, where the influence of the solar activity cycles increases sharply with increasing altitude [4.1.23]. In particular, carbon fiber structures are sensitive to atomic oxygen, but practical experience has shown that a normal satellite structure does not accumulate so much damage during its lifetime that the effect will become dangerous. For smaller parts, however, and for certain sensor structures in particular, the effect of atomic oxygen should not be neglected.

4.1.1.5 Inserts

For the fixation of payload boxes, service units of other subsystems, and all kind of secondary structures on the primary structure, a large number of screwing points are needed. In the case of a monolithically manufactured ("integral") structure it is possible to drill holes into the material, to tap suitable threads into them and to optionally apply helicoils. If the structure has been made in the traditional ("differential") manner, then cages for screw nuts or other thread-carrying elements can be welded, glued or riveted onto the structure.

Inserts in the usual sense of the word, however, are thread-carrying elements embedded in sandwich panels. They are by far the dominant fixation elements

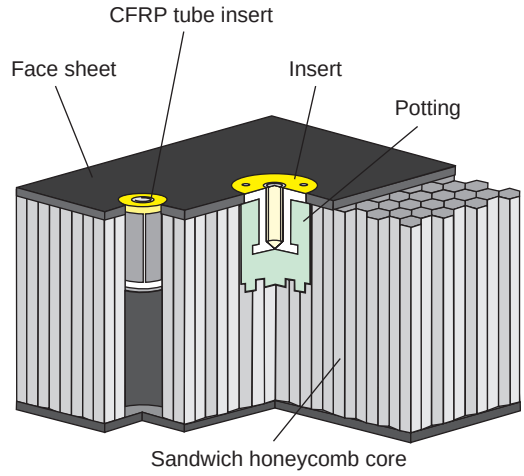


Figure 4.1.13: *Inserts in a sandwich panel.*

on spacecraft structures and can be classified into two basic forms: so-called **through-the-thickness inserts**, which go through the whole sandwich panel, so that the screw head and the unit to be fastened are on opposite sides; and the alternative version, so-called **blind inserts**, which involve unilateral inserts where the screw and the unit remain on the same side.

Both basic versions exist in various technical forms [4.1.24]. The standard form of a unilateral blind insert is embedded in the sandwich panel by the so-called **potting** procedure (see Figure 4.1.13, center). At first a borehole is drilled which is a bit deeper than the metallic insert element, but equal in diameter. Optionally the honeycomb cells touched by the drill are then closed at their remaining upper edge with some pasty epoxy resin so that, upon the following injection of liquid resin, this cannot flow into the "depths" of the sandwich. For comparatively thin sandwich panels this resin-saving step is of course not worthwhile, and the hole in the honeycomb core is then drilled through to the lower face sheet.

After implementation and alignment of the insert by means of an appropriate tool, liquid resin is injected through one of two holes in the insert cap, until the hollow space in the borehole around the central insert shank is completely filled and the resin pours out of the second hole. After hardening, each insert of this type provides a fastening opportunity for screws from M3 up to M6 and a load range of several kilonewtons, and contributes 4–7 g to the structural mass.

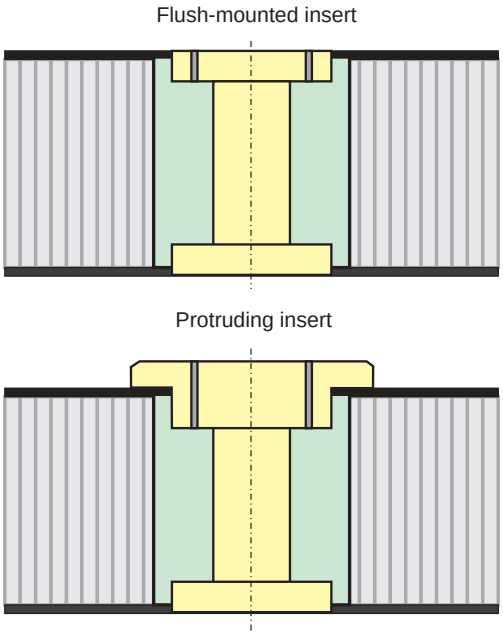


Figure 4.1.14: *Flush and protruding version of through-the-thickness inserts.*

In the left part of Figure 4.1.13 an alternative design is shown, a **CFRP tube insert**. For certain combinations of sandwich parameters it provides a better ratio between mass and load capability than does the conventional design. Here the borehole goes in any case through the whole thickness of the sandwich core, but its diameter is smaller than for the potting alternative. Drilling through the lower face sheet is optional. The borehole is then circumferentially wetted with some epoxy resin, and a foldable CFRP tube with extreme longitudinal stiffness, which is also wetted with resin on its outer side, is inserted into the hole. By means of a suitable tool this tube is subsequently unfolded to its full diameter, aligned, and positioned such that both tube ends fit exactly under the edges of the face sheet holes. Thus the whole length of the tube comes into adhesive contact with the surrounding honeycomb cell walls. After hardening, an aluminum insert which is in form-locking contact with a surrounding CFRP sleeve is glued into the tube [4.1.25].

Both inserts outlined in Figure 4.1.13 are so-called **flush-mounted inserts**: that is, their cap surface, on which the fastened units are tightly screwed, is flush with the surface of the sandwich face sheet. For most insert types, however, it is also possible to let the insert cap protrude above the face sheet (**protruding inserts**), as shown for comparison in Figure 4.1.14. The latter option allows for later collective abrasion of all inserts implemented in the same panel up to a common interface plane above the sandwich surface. Any local unevenness of the sandwich can thus be corrected.

All inserts must be implemented into the structure such that they cannot loosen during assembly and integration when the screws are tightened with the appropriate torque specified for the respective thread size. For fixation on space structures it is customary to use titanium screws (e.g., TiAl6V4) with a rolled thread and a hexagonal socket screw head. Since in particular the **random dynamic loads** occurring during the launch phase can easily lead to unscrewing, it is usual to additionally secure the insert threads with screw-locking elements, for instance bronze helicoils.

4.1.1.6 Manufacture and Integration

As already described in Section 4.1.1.3, the choice of the appropriate manufacturing technology for the body shell work on a primary spacecraft structure depends on both the construction principle and the material. This initial body shell work is in any case a "dirty" process, no matter whether it consists of the lamination of monolithic CFRP shells, the manufacturing of sandwich plates, or the milling, grinding and riveting of metallic sheet pieces.

The subsequent qualitative step from the "raw" body shell to the classified spacecraft structure, which then becomes the mechanical base for all further integration steps, involves three procedures:

1. **Cleaning:** The cleaning of the manufactured structure coincides with the transfer of all following activities to a **clean room**. It includes the removal of all loose dirt particles, swarfs, etc., by **purging** with compressed pure nitrogen, as well as the elimination of all surface impurities by washing with isopropyl. Fiber composite structures should have been cured once again at approximately 90 °C before this cleaning.
2. **Verification:** This involves exact weighing of the structure, or its constituent primary components,

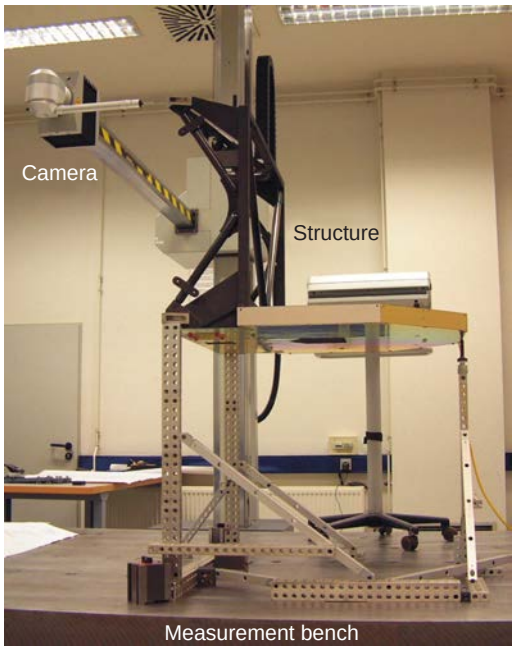


Figure 4.1.15: *Optical 3D gauging of a structure (Source: DLR).*

and precise optical verification of the geometry (Figure 4.1.15). Moreover, the position of the center of gravity (CoG) and, if required, the moment of inertia (MoI) in all three axes are experimentally determined.

3. **Configuration Control:** All structural parts are now subject to configuration control. The rule for their **labeling** is that all monolithically connected parts have one and the same label, while all removable parts have a separate label.

During the following integration the structure must be freely accessible, and its attitude must be adjustable in any desired position relative to the ground. Therefore the structure must be fastened in a rigid framework, usually made of steel, with the required geometrical degrees of freedom, the so-called **mechanical ground support equipment** (MGSE). It allows for the pivoting of the nascent spacecraft under normal terrestrial gravity and makes sure that lightweight structural parts which are designed for use in weightlessness are not accidentally overloaded during the ground operations. The mechanical interfaces for the fixation of the structure in the MGSE are therefore particularly important. Their number and load-carrying capability are selected such that even for large spacecraft the maneuver loads on the ground can never become a design-driving load case.

4.1.2 Secondary and Deployable Structures

Secondary structures differ from primary structure by not determining the "global" mechanical properties of the spacecraft as a whole. Rather they only govern the local behavior of the on-board systems or payloads that are fastened onto them. Nevertheless, their geometrical forms and sizes are such that they ought to be represented in the finite element model of the spacecraft in an identifiable manner, and their mechanical behavior ought to be explicitly analyzed. Only parts which are significantly smaller, such as compact boxes or cable clips, and which may be considered as **tertiary structures**, can be treated as "smeared mass" or as point mass. **Deployable structures** such as solar panels or booms are basically secondary structures.

4.1.2.1 Instrument Fixtures and Insulation

Instrument fixtures which can be regarded as secondary structures, due to either their size or their mechanical function (for instance, a nonnegligible cantilever beam), can be connected to the primary structure by means of inserts and screws, as can tertiary structures (e.g., boxes or clamps). Alternatively they can be monolithically connected to the primary structure, as shown in the example in Figure 4.1.16. In the latter case they are in formal terms, that is with respect to their **labeling**, a constituent part of the primary structure because they cannot be dismounted.

When fiber composite materials (preferably CFRP) are used, then such fixtures can easily and very precisely be aligned in the desired attitude and orientation, provided that a 3D CNC-manufactured building slip (or mold) is available, in which is embedded not only the primary structure, but also all interface planes and positions foreseen for the later attachment of the adjacent units. The previously manufactured adapter pieces between these unit interfaces and the primary

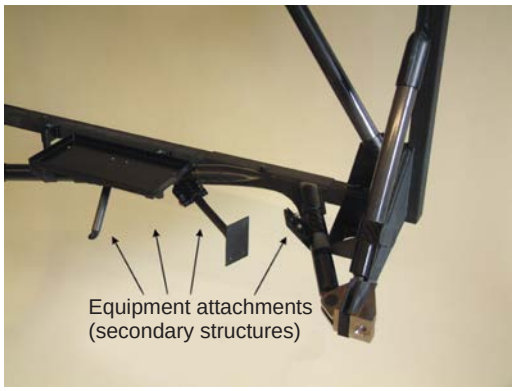


Figure 4.1.16: *Secondary structures for payload fixation on a load-carrying truss (Source: DLR).*

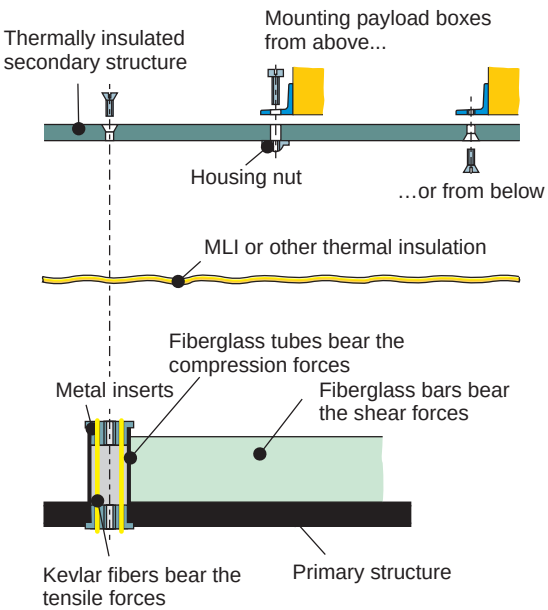


Figure 4.1.17: *Thermally insulated carrier structure.*

structure are "softly" glued in their positions so that they can harden while attached to the building slip. This insures that during later spacecraft integration the units will really fit without tension.

More ambitious are the carrier structures for payload or system units which must be held apart from the primary structure for thermal insulation reasons. As an example, Figure 4.1.17 shows a carrier

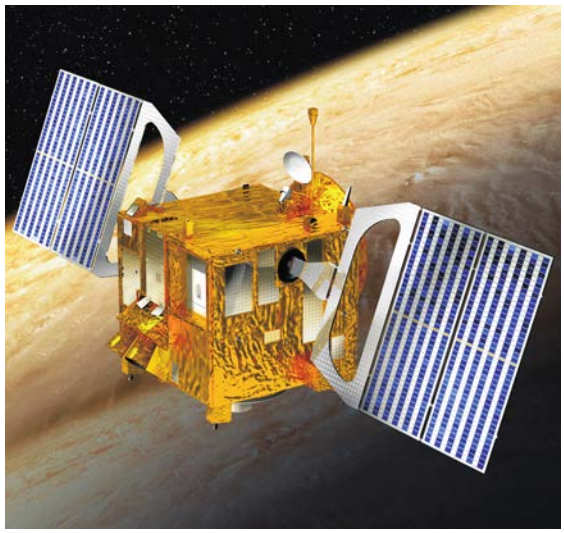


Figure 4.1.18: *Unfolded solar array panels of the Venus Express spacecraft (Source: ESA).*

configuration for electronic units which was developed for application in Mars projects such as the NetLander [4.1.20], [4.1.26].

The units are mounted on a carrier plate which is connected to the primary structure only by thin-walled **fiberglass composite** elements which penetrate through a thermal insulation layer so that thermal losses are minimized. In order to simultaneously bear the strong mechanical forces which may occur (e.g., during a landing shock), the load-carrying functions for tension, compression and shear forces are partitioned. The insert caps on the secondary side (warm) and on the primary side (cold) are sewed with **Kevlar** fiber rovings which, after impregnation with epoxy resin upon sewing, have been tightened and hardened.

4.1.2.2 Unfoldable Panels

Apart from the above-mentioned fixtures and adapters, the most frequently used secondary structures on spacecraft are unfoldable panels, in particular **solar array panels** for power generation.

The classical configuration of such solar panels is paradigmatically depicted in Figures 4.1.18 and 4.1.19. On each of the two opposite sides of the cuboid

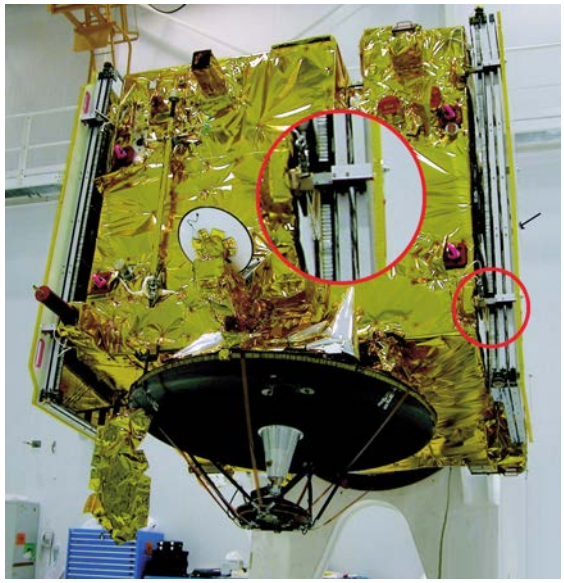


Figure 4.1.19: *Locked batch of solar panels on the Venus Express spacecraft (Source: ESA).*

primary structure of the ESA spacecraft Venus Express there are two rectangular solar panels. Their carrier structure is tapered on the inner side, toward the spacecraft, and attached to a so-called **bearing and power transfer assembly** (BAPTA) by means of which the whole solar array configuration can be actively tilted with respect to the position of the Sun.

The individual elements are connected by **spring-driven hinges** which open after unlocking in space and latch as soon as they have reached the unfolded state (see Section 4.1.5.3). Until just prior to launch, that is until the fairing is set on top of the launcher, it is usual to protect each batch of stowed solar panels by an additional **protective cover panel** in order to avoid accidental damage of the solar cells (see the enlarged display window in Figure 4.1.19).

Solar panels consist of relatively lightweight sandwich plates on which the solar cell arrays and the corresponding harness are glued. Under the **dynamic loading** during launch their mechanical behavior depends mainly on their firmly latching in the **launch locks** (see Section 4.1.5.1). The spring-driven hinges between the individual panels can be regarded as "soft" and almost neglected from a structural mechanics point of view.

As soon as the whole configuration has been deployed in space, however, the situation is completely different, particularly if there are three or more solar panels on each side (and not just two as in the above example of Venus Express, which cruises relatively close to the Sun). For instance, in the case of the Rosetta mission, which is bound for regions much more distant from the Sun, five large solar panels arranged one behind the other are needed on both sides of the spacecraft, achieving in all a "wingspan" of 32 m.

Even very small accelerations generated by flight maneuvers or minor disturbances can, under such circumstances, induce low-frequency vibrations in the order of 1 Hz. The stiffness (and correspondingly the first eigenfrequencies) as well as the damping properties of an unfolded solar panel configuration with latched spring hinges need in any case to be thoroughly analyzed. For the relevant analytical approach see [4.1.27].

4.1.2.3 Booms

In the context of space technology the term **boom** usually characterizes all kind of deployable cantilever arms by means of which instruments, sensors, antennas, etc., are positioned in a defined attitude outside the spacecraft.

If such a boom consists of rigid links with spring-driven hinges in between, then its structural mechanics are basically the one-dimensional variant of the batch of solar panels described above. The behavior of the stowed configuration during launch depends primarily on the quality and the position of the launch locks, while the dynamic behavior of the configuration in its unfolded state is determined by the mechanical properties of the boom links themselves, by the mass of the deployed payload (e.g., a magnetometer) and by the stiffness of the hinges.

One particular challenge is booms for the deployment of really large, ultralight structures such as solar sails, huge antennas or reflectors, or even solar cell arrays which are no longer deployed in the conventional manner by means of rigid elements, but by a flexible carrier which achieves its final contour only after having been unfolded in space. An example of these so-called **gossamer structures** is shown in Figure 4.1.20.

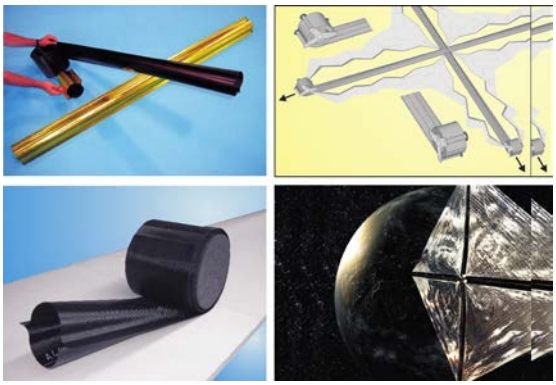


Figure 4.1.20: *Deployment concept for a solar sail spacecraft based on unrollable CFRP booms (Source: DLR).*

In the course of a joint technology study conducted by ESA, DLR and industrial partners, and aimed at the development of a future **solar sail** spacecraft which would be propelled only by the pressure of the solar radiation on a large reflective foil and could thus undertake long-term missions in the Solar System, ultralight booms had to be developed. It was mandatory to stow these booms during the integration and launch phases in a very small stowage volume [4.1.28].

As depicted in Figure 4.1.20, each boom consists of two long, unrollable, very thin-walled curved CFRP shells, the half-circled profiles of which are the mirror image of their respective counterpart. In the absence of external forces, these shells unfold under the influence of internal stresses that have been "frozen" during hardening, and form a long tube with two sideward flanges. Such a tube can be compressed to a flat ribbon and rolled up on a spool like a tape measure. Upon unwinding, the stored deformation energy initiates the unfolding. By means of a suitable mechanism several booms can be radially deployed out of a central body and simultaneously unfold a thin sail foil or another ultralight device [4.1.29], [4.1.30], [4.1.31].

Since these booms are very thin walled, their load-bearing capability depends primarily on their buckling stiffness, for which a laminate based on a combination of 0° and 45° fiber layers is very advantageous. Moreover, this lay-up sequence helps to minimize the deformation of the boom under unilateral thermal irradiation in space. Very low coefficients of thermal expansion (CTEs) in the longitudinal direction are decisive. The specific weight of the boom is in the order of 100 g/m.

The dominating thermal effect on such booms is heat transfer by radiation, because their surface is huge in relation to their mass. Therefore future applications will decisively depend on a suitable surface coating which is anyway necessary as a protective shield against **degradation**. The use of coated Kapton foil is already proven, at least on shorter boom sections. Also, the deployment concept as such was successfully tested in the course of a large-scale on-ground demonstration in 1999, where a solar sail with four booms each 14 m long was unfolded [4.1.30], [4.1.31].

4.1.3 Structural Analysis

4.1.3.1 Finite Element Models and Modal Models

Considering the complexity of spacecraft structures it is evident that their mechanical behavior cannot be comprehensively and accurately described by analytical formulas alone. At best some global properties can be approximately estimated by "hand calculations" (which is anyhow valuable in Phase A studies).

Therefore the structural mechanics calculations are today performed by means of the **finite element method** (FEM) as the state of the art in engineering sciences. Using the FEM means representing the real structure by a discrete mathematical model which consists of a large number of geometrically simple finite elements that are connected to each other at defined nodal points, thereby forming a mesh. In the case of the simplest (linear) finite element types these **nodes** are identical to the corner points of the element. The elements can be one, two or three dimensional. Corresponding to the mechanical properties assigned to them, they can be classified as **rod elements**, **beam elements**, **shell elements**, **solid elements**, and many other types.

Any attempt to thoroughly describe here all these elements and the idealizations upon which they are based would be far beyond the scope of this handbook. Therefore reference is made to the comprehensive literature on this topic, in particular to the excellent monograph by Bathe [4.1.32]. Moreover, for each of

the common FEM software packages (e.g., **NASTRAN**, **ANSYS**, **ABAQUS, ASKA**) there is a detailed specific description of all elements available in the respective software.

The generation of a suitable FE model can be facilitated if a **CAD model** already exists from which it can be derived. Thus it may not be necessary to model the whole geometry of the structure from scratch. However, the designer's view is sometimes different from the view of the engineer responsible for the structural analysis. For instance, a thin-walled shell structure may be ideally represented in the FE model by shell elements, based on which the numerical approach is two dimensional, namely superposition of a plate-bending stiffness and a membrane stiffness. The CAD designer, however, may prefer to work with "solids," that is with volume elements, because the generation of technical drawings and the manufacturing requirements necessitate this approach. In any case the conversion of a CAD model into a FE model must be examined very critically.

It is obvious that the accuracy achievable with FEM calculations depends on the mesh, that is on the resolution of details by a number of elements. Since a very fine mesh leads to strongly increased modeling and calculation effort, it is restricted to the vicinity of critical interfaces, essential load paths and other important areas. On the other hand, for parts of the structure where "nothing happens," for instance areas in the interior of large plates, a coarse mesh is sufficient.

Of course the FE model of a spacecraft structure must always include the mounted system and payload units. A mechanical analysis of the "naked" structure would be senseless. Relatively light and small units (**tertiary structures**) can be represented by the idealized assumption that their mass is concentrated at their CoG which is connected to certain nodes on the structural mesh (close to the real interface points) by rigid, but mass-less connections, so-called **rigid beam elements** (RBEs), see Figure 4.1.22 below.

Relatively heavy and critical units, on the other hand, need a detailed representation in the FE model. In space projects, the suppliers of such a unit may be obliged to provide their own FE submodel which can subsequently be incorporated into the overall FE model of the spacecraft.

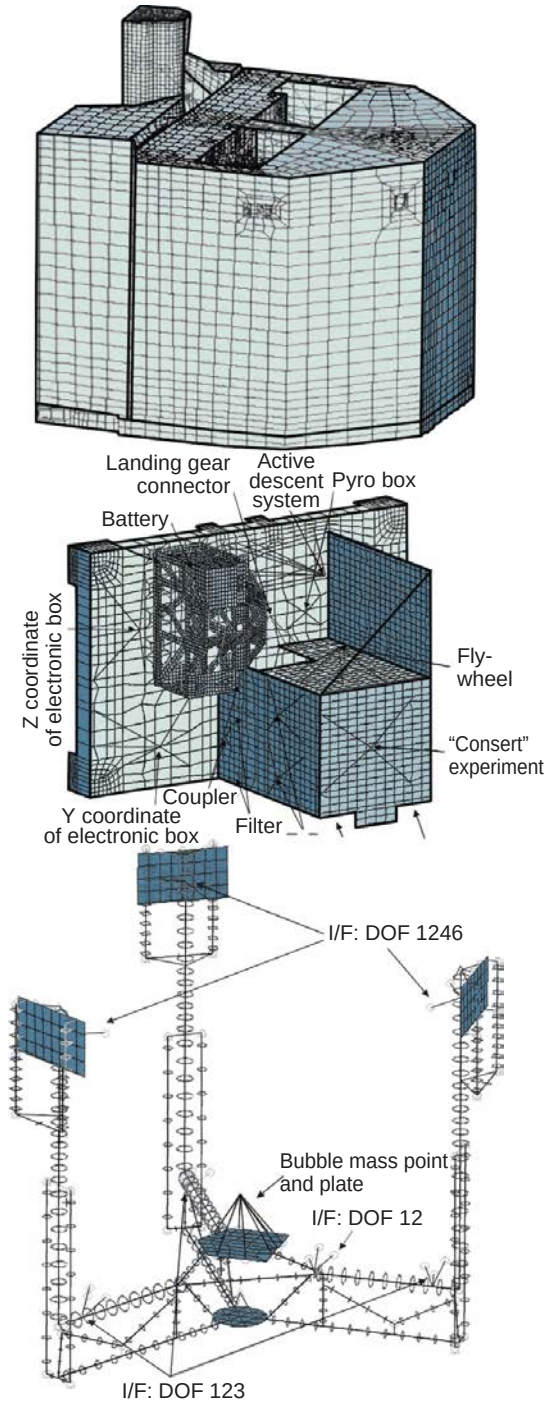


Figure 4.1.21: *FEM of the Rosetta lander Philae (detail).*

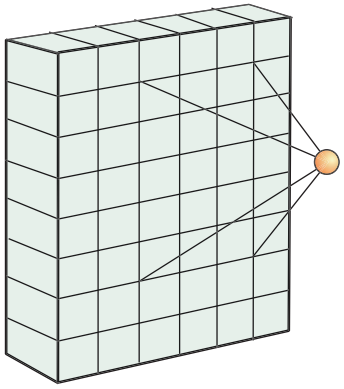


Figure 4.1.22: *FEM idealization of small units.*

An example is the battery rack on the instrument carrier of the Rosetta lander, as depicted in Figure 4.1.21 (center). Originally the battery had been idealized in the same simplified manner as shown in Figure 4.1.22, until a significant mass increase led to a negative **margin of safety** (**MoS**, see Section 4.1.3.3) at several battery fixation points and to critical dynamic behavior. Consequently the supplier company was put in charge of the delivery of an FE submodel which, after incorporation into the superior model, allowed the necessary constructive improvements to be determined.

Masses which are difficult to localize or do not contribute anything to the stiffness of the structure are treated as **nonstructural mass**, for instance foils, MLI, harnesses, or solar cell arrays. They are "smeared" on the adjacent structural parts like a coating.

For practical calculations the FE model is fed with the relevant material parameters, such as the different elasticity and shear moduli, so that all properties are reduced to values assigned to the nodes of the mesh. The result is a system of equations describing the displacements of the nodes and the forces between them. Such a system of equations may be quite comprehensive, but the underlying physics is in the simplest case a linear stress–strain relation, that is Hooke's law. For dynamic analyses there are supplementary terms describing the inertial and damping properties. Of crucial importance are the boundary constraints which define whether the structure is somewhere fixed, whether it can freely vibrate, or which other conditions may apply.

Before the FE model is actually used it is commonly subjected to a **model verification**. At first a so-called "mass properties check" is performed; that is, mass, CoG, and MoI with respect to the main axes of inertia are determined so that their plausibility can be checked. In NASTRAN, for instance, the next steps are a "rigid body check," a "residual strain energy check" and a "static load check." All these checks are basically tests for the self-consistency of the model.

Apart from the FEM there are in principle other discrete mathematical modeling methods, in particular the **finite difference method** (FDM) and the **boundary element method** (BEM) [4.1.33]. For the purposes of structural mechanics, however, they are less suitable than the FEM. The FDM is very applicable for the calculation of transient temperature distributions, and for this function it is also established in space technology (see Section 4.3). The BEM is advantageous for some special problems, for instance the determination of stress concentrations on crack tips or notch tips.

The above-mentioned FE models are also defined as physical FE models because they represent (with more or less accuracy) the real physical configuration of the structure. However, if the structure is to be integrated into a superior unit, then the details of its behavior are meaningless for the structural analysis of the larger unit. Of interest is only how the implemented small unit as a whole acts on the interfaces to the larger superior unit.

For static analyses the physical FE model of the embedded unit can be replaced by a **reduced stiffness matrix**. For dynamic analyses it is then sufficient to determine the global "response function" of the embedded component under a given excitation. The respective models are called **modal models**. They are derived from the physical FE model by special methods such as the **Craig–Bampton method**.

Responsibility for generation of the FE model and for the numerical analyses performed with it is usually assigned to the subsystem "structure" (i.e., to the institution or company in charge of providing the structure), even though the FE model always depicts the spacecraft as a whole.

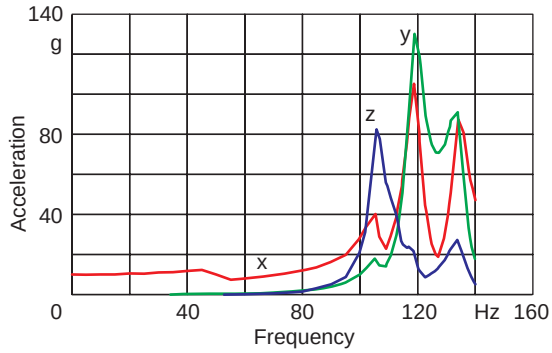


Figure 4.1.23: *Spectral structural response (x-excitation).*

4.1.3.2 Resonance Behavior and Eigenmodes

The **resonance behavior** of the structure, that is the frequency spectrum of the **eigenmodes** and the amplitude obtained at a given excitation (the so-called "structural response") is at first numerically determined by means of the FE model.

For the excitations in each of the three spatial directions (x, y, z) the "responses" are calculated also for all three directions, as shown in Figure 4.1.23, so that in total $3 \times 3 = 9$ response curves are obtained. Then it must be verified that the first global **eigenfrequencies** lie beyond a specified limit (for instance, 100 Hz) in order to exclude critical resonances between the excitations generated by the rocket engines and the spacecraft. An eigenfrequency or eigenmode is "global" if a substantial portion of the **modal mass** is involved in the vibration. If, on the other hand, only a few percent of the modal mass participate in a vibration mode, then this mode is "local" and in most cases induced by some identifiable subunit. The relative amount of involved modal mass is determined automatically by the FE program.

The real spectral response functions of the structure, or of the whole, fully integrated spacecraft, are repeatedly determined in the course of the **structural tests** within the AIV program. Their experimental determination is achieved by **low-level sine vibration tests** that are performed before, between and after all qualification tests under higher loads, in order to verify the enduring integrity of the spacecraft (see Section 4.1.4). These results are used to adapt the FE model to reality. The model parameters are iteratively modified until the results gained with the FE model correspond exactly to the experimental ones (**model updating**).

Table 4.1.2: *Common safety factors (example).*

Materials	Plasticity (yield)	Fracture load (ultimate)	Buckling load
Conventional	1.25	1.5	2.0
Unconventional	1.5	2.0	2.0
Inserts and joints	1.5	2.0	n/a

Additional safety factor	With special test	Without special test
Curves, structure inserts	1.1	2.0
Strain on the honeycomb core	1.65	Forbidden
Breaking load (at ultimate/yield < 1.2)	1.0	1.7

4.1.3.3 Strength Proof and Margins of Safety

After the analysis of eigenfrequencies and eigenmodes, the **strength proof** of the structure becomes the focal point of the calculations. For all structural components and all relevant connections and interfaces (inserts, struts, etc.) it has to be proven that the calculated allowable loads at the respective location are higher than the maximum forces which may occur with the given design, where these "design loads" are still to be multiplied by one or more safety factors (see Table 4.1.2).

Consequently, the defined ratio should be greater than one. But since a change of sign appears even more striking than does the fluctuation of a number around one, the so-called **margin of safety** (MoS) is pragmatically defined as

$$MoS = \frac{\text{allowable load}}{\text{design load} \times \text{safety factors}} - 1$$

This value is determined for all relevant joints and interfaces on the structure, in particular for all inserts onto which assembly parts of nonnegligible size and mass are fastened. If the MoS value becomes negative this is immediately obvious so that corrective measures can be initiated.

Table 4.1.3: *Margins of safety at some of the instrument feet (example: battery development of the Rosetta lander).*

Unit/component	Fastening point no.	Max. load (strain, compression)	Max. load (pressure)	MoS	Measures
Battery with bracket	1	1237	587	–0.068	
	2	1237	587	–0.068	
(originally fastened at 6 points with the inserts, weight increase > 5 kg)	3	1111	593	0.024	Negative MoS at 4 out of 6 interfaces leads to an increase of fastening points and design improvements
	4	1409	713	–0.187	
	5	1409	713	–0.187	
	6	1111	593	0.024	

The example shown in Table 4.1.3 refers to the battery fixture on the Rosetta lander that was discussed above. Due to a significant mass increase of the battery, the MoS values calculated for some of the feet of the fixture became negative. This indication led to the demand for a FE submodel from the battery supplier in order to allow for more exact calculations, and subsequently to some significant design improvements.

4.1.3.4 Modeling of Critical Interfaces

There are structural interfaces the mechanical properties of which can decisively govern the whole behavior of the attached structure. In particular, the fixations of a spacecraft structure on a superior unit, for instance the interface between satellite and launch adapter, subsatellite and main satellite, lander and orbiter, etc., belong to this category.

In these cases it is not sufficient to rely on a very fine mesh alone, nor is it admissible to simply presume a completely rigid connection. Rather, a spring constant must be introduced which cannot be reliably validated until the structural tests have been evaluated (Figure 4.1.24). Only an empirical comparison of the movements of the spacecraft and the rigid body movements of its base mounted on the shaker table, which are measured by special accelerometers (so-called "pilots"), gives evidence of these model parameters. Therefore, it is essential to perform the structural tests with a realistic carrier structure similar to the real flight configuration and not to rigidly fasten the spacecraft structure onto the shaker table itself (Figure 4.1.25).

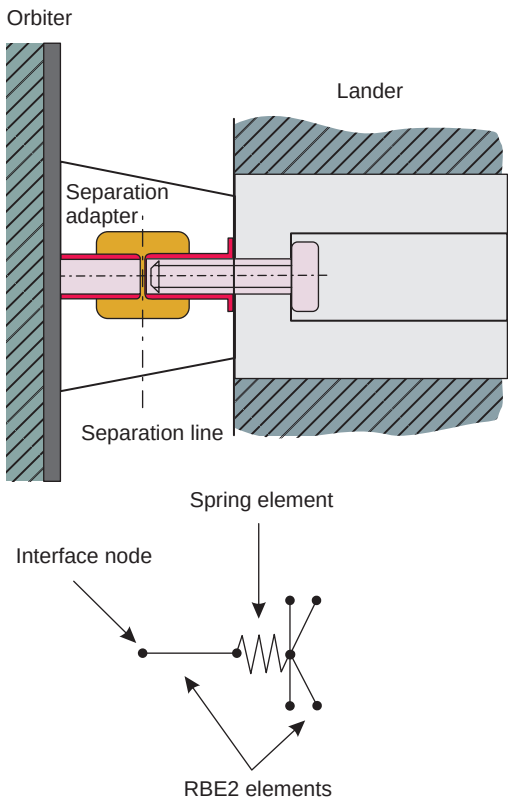


Figure 4.1.24: *FE modeling of the interface nodes between Rosetta orbiter and lander.*

The same is true for lids and locks on the structure itself which have a firm abutment on one side, but are held toward the other side only by a down-holding mechanism. In order to avoid unrealistic modeling



Figure 4.1.25: *Structural tests on the Rosetta lander (Source: IABG).*

it is recommended that the real elastic properties are clarified by means of breadboard tests on simplified test hardware.

4.1.4 Qualification of the Spacecraft Structure

The structure of a spacecraft, usually followed by the thermal control system, belongs to those subsystems which need to be qualified as early as possible in the course of the project. The reason is obvious: if the qualification is not successful, then the need may arise to redesign the whole configuration from scratch, with serious consequences for the development schedule and costs, and also for the other subsystems and the payload. Compared to the structure, it is much easier to isolate and finally overcome the consequences of any design errors or weak points in other subsystems, for instance in the electronic ones, probably by "delta qualifications."

The qualification is normally performed by means of the **structural–thermal model** (STM), the first of the "official" qualification models, in most cases at the beginning of Phase C/D. An STM is already completely subject to the whole set of configuration control rules, other than any proceeding **breadboard models**.

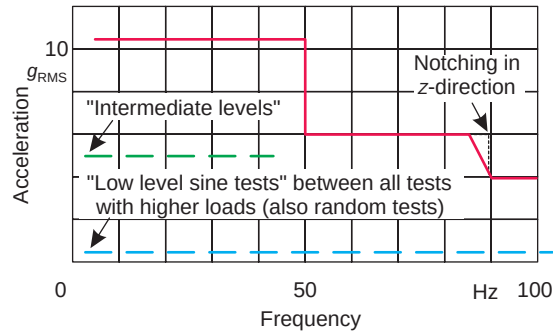


Figure 4.1.26: *Load levels in structural tests.*

The STM must represent the spacecraft in the fully integrated state; that is, all units on-board must be represented by models ("mass dummies") which are realistic with respect to their mass, CoG, MoI, external geometry, surface properties and mechanical interfaces. (For the thermal tests also their heat exchange with the environment must be simulated realistically.)

Figure 4.1.26 demonstrates the test philosophy, taking the sine vibration tests as an example. The STM is separately tested in each of the three directions (x, y, z). At first a so-called **low-level sine test** with a very low excitation amplitude (probably 0.5g) is performed. In this test the frequency is slowly (e.g., at 2 octaves per minute) swept from almost zero up to 2000 Hz, so that the resonance curve can be determined. Subsequently the excitation amplitude is stepwise increased up to the **qualification level** (**Q-level**); however, these real loading test runs are usually limited to frequencies up to 100 Hz.

Between all these intermediate-level or Q-level load tests, low-level sine tests are repeatedly performed in order to verify the enduring integrity of the structure. A frequency shift of the resonance peaks between two low-level sine test curves recorded prior to and after a "real load" test would indicate internal damage, that is a progressing "softness" of the structure. The frequency response curve (resonance curve) measured in the final low-level sine test for each axis then serves as the basis for the model updating of the FE model.

For **sine load** tests the required excitation amplitudes in the directions of all three axes are predefined

as effective values (root mean square, RMS) versus frequency with the acceleration of gravity as the unit, that is in g_{RMS} (see Figure 4.1.26). **Quasi-static loads** are measured in multiples of g as well.

For a **random load** and **acoustic load** the required excitation in each axis is predefined as a **power spectral density** (PSD), which is pragmatically measured in g^2/Hz. The same approach is used for **shock loads** [4.1.34]. At least the random load tests are accompanied by preceding and succeeding low-level sine tests, as described above.

In case of critical resonance magnifications it is allowed under certain circumstances to reduce the excitation amplitude on the Q-level in limited frequency intervals. The originally prescribed acceleration curve (scaled in g versus Hz) or PSD curve (scaled in g^2/Hz versus Hz) then gets a notch. Therefore, this procedure is quite pictorially called **notching**. It needs to be formally approved in each individual case by a "Test Review Board" appointed by the project management.

The Q-level is usually 1.5 times higher than the so-called **acceptance level** (**A-level**), which still covers sufficiently safely the dynamic loads that can be really expected during launch.

Other than the STM, the **flight model** (FM) is only tested on the A-level, but apart from that, the test procedures for the FM are the same. The FM is mounted on the shaker table in all three axis directions (x, y, z) one after another (see Figure 4.1.25) and tested under the prescribed sine and random loads, repeatedly, with low-level sine tests in between. Acoustic tests and shock tests are optional, depending on the specific project requirements. For the FM they are often omitted in practice.

4.1.5 Mechanisms

The following considerations are confined to **mechanisms** which are directly connected with the structural mechanics of the spacecraft, for instance by serving to separate different segments from each other, to lock and unlock, or to deploy and actively position substantial spacecraft parts. Mechanisms in the interior of payloads (e.g., inside scientific instruments), however, are not considered here.

A basic requirement for spaceborne mechanisms is high autonomy or reliable functioning without any human assistance. Only in exceptional cases, for instance on the International Space Station (ISS) or on-board the Space Shuttle, astronauts are present who could help a jamming mechanism along. In most other cases, however, there is no chance to intervene, and a failing mechanism can well be hazardous for the whole mission.

While for many mechanisms designed for nonrecurring use the holding time from launch to release is short (minutes, hours, or a couple of days at the most), tilting or positioning drives must often remain continuously operable under space conditions for many years. In the case of extraterrestrial missions the criterion of long-term reliability is also mandatory for separation, unlocking and spring mechanisms, which must immediately function at the target of the mission even after a long cruise flight and under extreme environmental conditions.

4.1.5.1 Pyromechanisms

Pyromechanical separation mechanisms serve for the fast and reliable separation of different spacecraft segments from each other, or for the transection of cables and ropes. In the case of rigid connections they are referred to as **pyrobolts**, while for the transection of cables and ropes the terms **cablecutter** or **pyroknife** are customary. All of them are sometimes summarized by the term **ordnance devices**.

The constructive diversity of such mechanisms, which are commercially available, is rather large. In all cases a small explosive charge is ignited in a reaction chamber so that the explosion pressure can act via a piston (or the equivalent of a piston) on a cutting tool, by which a rigid bolt or a cable can be transected.

`Since an explosive charge generates very large forces or pressures, the mechanism can be quite small for most applications in space technology. The largest pyrobolts are generally needed to separate the different stages of the launcher.

For the latter purpose also the second characteristic of pyrobolts (apart from the high forces released) becomes decisive, namely the short duration of the reaction. Rocket stages must be separated in one fell swoop, and if they are connected by several pyrobolts

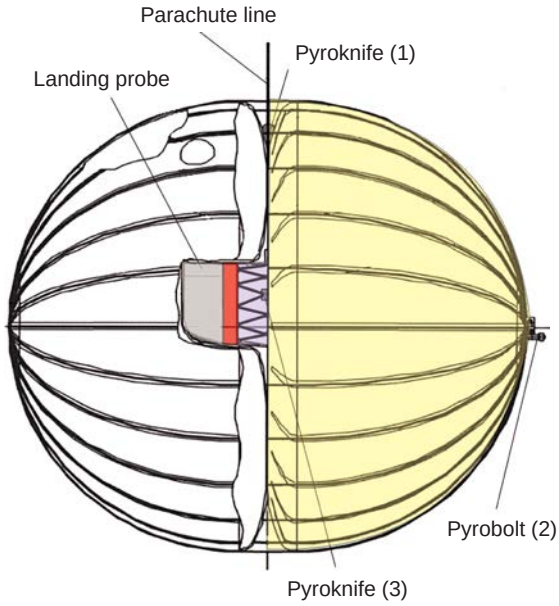


Figure 4.1.27: *Pyromechanical airbag separation release of a Mars lander (Source: CNES).*

these must be fired synchronously. This synchrony is only realistically feasible by means of pyromechanical devices.

Upon ignition **shock waves** in the structure are inevitably generated which, in the frequency range of several kilohertz, can reach an amplitude of several thousand *g*. Typical of the "far field" at some distance from the pyrodevice is a rise in amplitude up to a "knee frequency" of approximately 5 kHz with a subsequent amplitude decay [4.1.34]. Within the analysis and qualification of the spacecraft these shock waves must necessarily be considered. In particular, sensitive instruments can easily be damaged by shock waves and may need a damping bearing [4.1.35].

Figure 4.1.27 shows the deployment concept of an airbag designed for the landing on Mars (NetLander project [4.1.20]). The concept was derived from a previous development of the Babakin Company for the Russian Mars-96 mission and was based solely on pyromechanical devices.

Both calottes of the airbag which surround the landing probe are laced up to a single sphere. After transection of the parachute line by means of a pyroknife on the upper side and touchdown on the Martian surface, a pyrobolt releases all the ropes surrounding the sphere in one fell swoop, which initiates the immediate bellying of the calottes. The subsequent pyromechanical transection of ropes and straps in the interior, which circumferentially surround the landing probe, accomplishes the landing process.

Apart from separation mechanisms in the literal sense, there are also pyromechanical **unlocking mechanisms**, in particular **pinpullers**, since a large portion of all launch locks of panels, booms, antennas, etc., is based on pins which hold a spring-driven or motor-driven deployment mechanism by form-locking contact. Here it is decisive to pull the locking pin reliably and without jamming of its counter-bearing so that the deployment can occur unimpeded.

If the unlocking must occur at several locations synchronously, then pyromechanically driven pinpullers are often preferred in order to avoid having the deploying structure free on one side while it is still locked on the other side, which might lead to jamming or hazardous dangling movements.

The external shape of pyromechanical pinpullers is in most cases quite similar to that of pinpullers with a nonexplosive drive (Figure 4.1.28). The reaction takes place in a cylindrical chamber, after which a bolt is abruptly pulled behind or pressed out of a flange plate.

All pyromechanical devices are subject to particular safety regulations and may be armed only immediately prior to launch. On satellites and space probes it is often only allowed to install them in this phase, replacing the nonhazardous placeholders. In any case they require special access conditions.

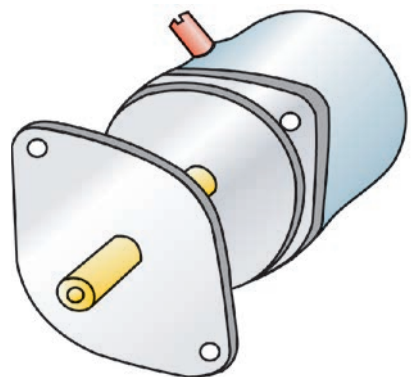


Figure 4.1.28: *Pinpuller (schematic).*

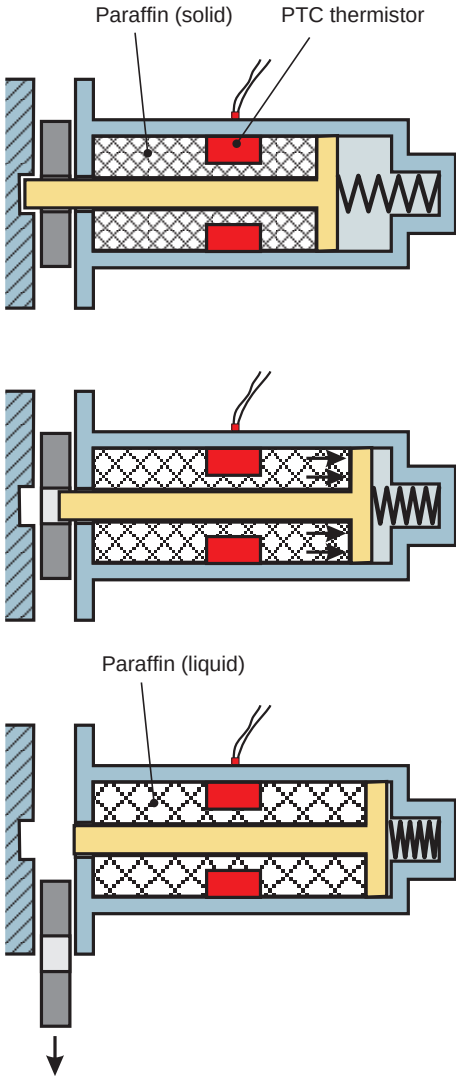


Figure 4.1.29: *Wax motors as pinpuller (schematic).*

4.1.5.2 Nonexplosive Actuators and Launch Locks

If a short reaction time, or synchrony in the case of several release locations, is not a prime criterion, then nonexplosive actuators can also be used. Similar to pyromechanically driven devices but much slower, a chemical reaction is ignited in their reaction chamber, or a phase transition is initiated, which pressurizes a piston or something equivalent. By means of this pressure, bolts or ropes can be transected, but also locking pins can be pulled. Shock loads are avoided.

For instance, the four separation adapters of the Rosetta lander schematically depicted in Figure 4.1.24 were so-called **nonexplosive actuators** (**NEAs**) which had to transect screw connections with M9 thread. The transection was performed shortly after launch in order to avoid the risk of thermoelastic stresses which could, after 10 years of cruise flight in space, possibly lead to jamming. Nevertheless this transection of the four main load paths between orbiter and lander was not the same as an immediate physical separation. Both spacecraft units remained connected to each other by a light holding mechanism (cruise latch).

One of the most reliable types of pinpullers with a nonexplosive drive is the so-called **wax motor** (see Figure 4.1.29). A cylindrical chamber contains solid paraffin in which a heating element is embedded, mostly a PTC thermistor, optionally in duplicate. Upon melting, the paraffin expands and presses with a high force against a piston which acts in turn on the actuator pin. Depending on the construction, the pin can be either retracted or pushed forward. Suitable mechanical deflections allow for all kinds of unlocking movements. Of course there are also electromagnetically driven pinpullers. However, with respect to the generated force in relation to the required electric current, the size, and safety against jamming, wax motors are clearly superior as long as the velocity is uncritical.

Pinpullers and other types of launch locks are today commercial available as **commercial off the shelf** (COTS) products in a broad variety of forms, and need no longer be developed for a specific mission.

4.1.5.3 Spring Mechanisms

The above-mentioned pinpullers are very often symbiotically connected to **spring mechanisms**, where they unlock a prestressed compression spring or torsion spring and thus permit deployment of the attached unit (boom, panel, antenna, etc.).

The **torsion-spring-driven hinge** depicted in Figure 4.1.30 is a typical example. It was developed by the Pfeil Trawid Company [4.1.36] for a Mars landing probe (NetLander project [4.1.20]), for the deployment of its solar array petals that were stacked upon each other. The torsion spring has a diameter of 21.8 mm, a length of 32 mm and generates a maximum torque of 2.1 N m. The whole hinge weighs 68 g.

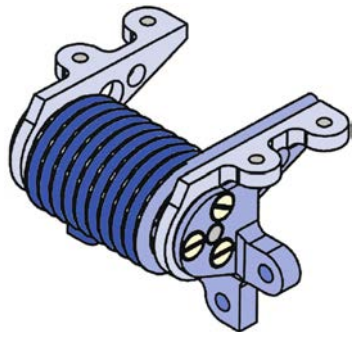


Figure 4.1.30: *Typical torsion-spring-driven hinge.*

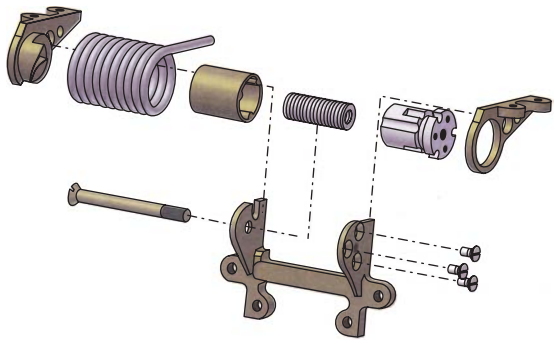


Figure 4.1.31: *Internal view of the torsion-spring-driven hinge.*

Since the moved component (in this example the unfolding solar array petal that is fastened on the upper flange) is accelerated during the rotational movement of the spring, it could hit the stopper with maximum velocity, that is with maximum angular momentum, which would result in strong oscillations, probably even damage. Therefore the torsion-spring-driven hinge contains in its interior a **torsional brake** which becomes progressively effective with increasing opening angle.

As shown in the exploded view (Figure 4.1.31) the opening of the torsion spring is accompanied by the co-rotation of a cam-guided cylinder in a guiding sleeve which compresses a **Belleville spring washer** such that the rotational movement is effectively damped until the final position (latching position) is reached. Spring hinges of this type are well proven in space technology.

Since in many applications several similar deployment steps must be taken consecutively (for instance, in the case of sequential unfolding of solar panels) the spring mechanisms and their unlocking devices must be effectively switchable in series.

If the moving parts are somewhat distant from each other then each individual spring-driven hinge will be released by a pinpuller (driven pyromechanically, electromagnetically or by a wax motor) and the individual steps of the release sequence will be controlled electrically. For instance, upon complete unfolding of a panel or boom link, the unlocking of the next one can be triggered by an end switch.

If the moving parts are in close contact, then the control can also be purely mechanical. In the above-mentioned example of the sequential radial deployment of the solar array petals of a small Mars landing probe, each petal bears a spigot on the inner side of the flange of the torsion-spring-driven hinge which locks the subjacent petal in the stowed configuration. As soon as the hinge has opened up to a certain angle, the subjacent petal is released (Figure 4.1.32).

Spring-driven hinges can also be combined with electrical hinge drives. The stack of solar array petals developed for the above-mentioned NetLander project was covered by a much more massive outer lid which served not only as solar array carrier, but also as bearing area for a possibly necessary uprighting movement (see Figure 4.1.11). The hinge drive for this outer lid had of course to be electrical, because the necessary uprighting torque was 25 times higher than the torque required for the lid opening only. However, the latter case was the nominal one, because the nominal landing attitude was upright.

In order not to make mission success in the nominal case (upright position) conditional on the functioning of the motor-driven hinge (i.e., to avoid a so-called **single point failure**) the hinge axis was additionally equipped with a torsion spring drive which could at least open the lid in the case of a motor failure (see Figure 4.1.33).

4.1.5.4 Electric Motors and Drives

Figure 4.1.33 shows exemplarily the above-described combination of a motor-driven and a spring-driven hinge [4.1.36], see also Figure 4.1.11. The necessary strong gear reduction between the relatively small **electric motor** (on the right side of the figure) and

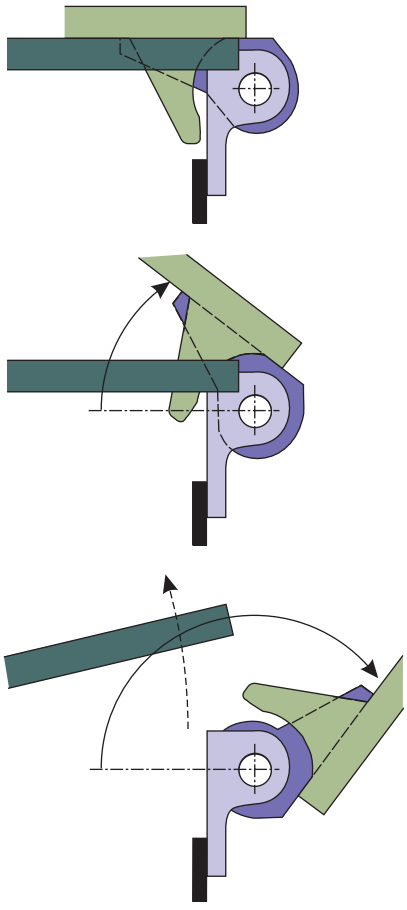


Figure 4.1.32: *Sequential unlocking of petals with spring-driven hinges.*

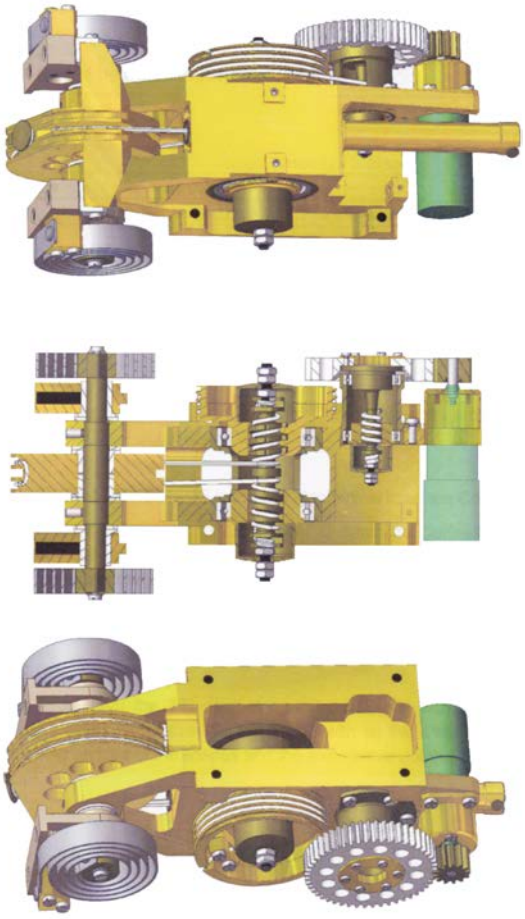


Figure 4.1.33: *Motor-driven winch-based gear with additional torsion springs (Source: Pfeil Trawid).*

the hinge (left side, with attached springs) is achieved in four steps.

First, the motor itself carries a **planetary gear** which ends in a gearwheel made of titanium. It transfers its torsional moment to a larger gearwheel made of PEEK in the second step of the gear reduction. Thanks to the material combination titanium/PEEK, friction problems between the two gearwheels are practically excluded. The third and fourth stages of the gear reduction are winches which spool a 1.6 mm thick Dyneema string. In this application, where in the final stage large torques are required but short pulling distances are sufficient (because the hinge axis rotates only once by 180°), **winches** are much more advantageous than gearwheel mechanisms. Otherwise, one would need either more axes or a much larger gearwheel on each stage.

It is evident that such winch-based gears are only suited for one-way mechanisms. For tilting and positioning drives which must work continuously and in both directions, gearwheel drives or worm gears are used, which differ from comparable terrestrial ones only by their particularly lightweight design and by the specific tribological aspects which must be considered in space (see Section 4.1.5.7).

The electric motors which are commonly used in space mechanisms can be classified in four main categories [4.1.37]:

1. DC motors with brush contact for the generation of substantial torques. They basically consist of a stator with permanent magnets, a rotor with a commutator and a commutator ring with brushes. This very conventional configuration is suitable as a "workhorse" for a very limited duration, because brush contacts under space conditions become very quickly unusable.
2. Brushless DC motors which can only rotate through a very limited angle but are extremely simple because they need no commutator ring. They can be used, for example, for swiveling and pointing movements for which no gear (such as the one described above) is needed, because the required torque is sufficiently low.
3. Brushless DC motors which need no commutator because the coils are part of the nonrotating motor component. Although this construction is known as a "continuous-rotation, brushless DC motor" (**BDC motor**), it is strictly speaking working with alternating current because it partitions the applied direct current into phases according to the angular position of the axis [4.1.37]. The big advantage of this construction is its relative insensitivity to the environmental conditions in space. Brushless DC motors are therefore widespread in space technology, and are mostly already manufactured as an integral combination with a planetary gear.
4. **Step motors**, as used also in terrestrial applications, in order to accurately control the axial movement in discrete steps (see also [4.1.40]).

Small displacements can also be generated and controlled by **adaptive structural elements**. Actuators and sensors which are embedded like an implant in the primary spacecraft structure or in the respective secondary structure (truss, fixture, etc.) can be electrically contacted and integrated into an electronic control circuit, which allows for highly precise attitude and shape variations, both statically and dynamically (**smart structures**).

Here the broadest diversity can be achieved with **piezoceramic materials**, for instance barium titanate or lead zirconate titanate (PZT), which are mostly manufactured in the form of thin plates but can be stacked to so-called stack actuators. Depending on the axial orientation of the (piezo)crystal relative to the macroscopic dimensions of the actuator, and as a function of the applied voltage, elongations, contractions and deflections can be generated. Other adaptive materials, for instance **shape memory alloys**, are at the present time still less important than piezoceramics because they can only react slowly to electrical signals and are therefore unsuited to dynamical applications (e.g., for active vibration damping).

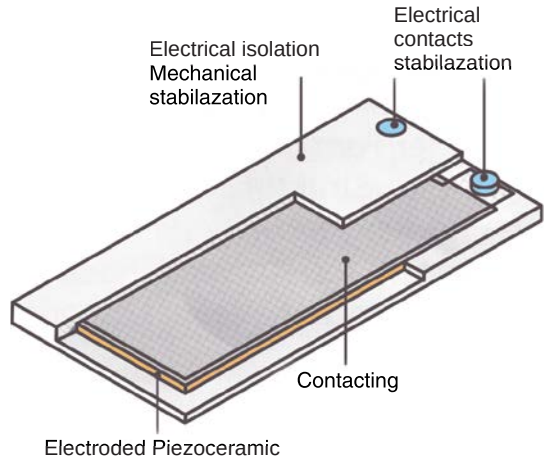


Figure 4.1.34: *Piezoactuator embedded in CFRP.*

Figure 4.1.34 shows schematically the embedment of a piezoceramic actuator in a **CFRP laminate** [4.1.38]. Only in this rigidly embedded form can the extremely brittle piezoceramic material (in this case PZT) be prevented from fracture and become operable at all. The displacements generated with the actuator are transferred to the overall structure via the elastic properties of the embedding structure.

An exemplary space application is shown in Figure 4.1.35. Within a joint project between DLR and DaimlerChrysler Aerospace Dornier a satellite mirror with piezoceramic actuators for adaptive shape control was developed and successfully tested [4.1.39].

Optical mirrors for astronomical applications (space telescopes, interferometry) or for use as laser reflectors must have extremely precise contours. The suitability of passive carrier structures for this purpose is limited because the occurrence of disturbances which cannot be compensated passively can never be excluded (e.g., those caused by different thermal expansion). In the foreseeable future the mission

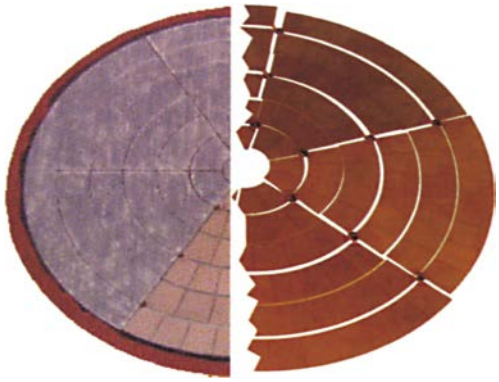


Figure 4.1.35: *Adaptive shape control of the carrier structure of a satellite mirror.*

requirements for such mirrors will make active shape control indispensable, and it is obvious that the installation of numerous motors and drives on the spacecraft for this purpose would be much too complex. Indeed they are not needed because the required displacement amplitudes are very small.

For the mirror carrier structure depicted in Figure 4.1.35 with a diameter of 600 mm, 180 actuators were foreseen, distributed on 36 ring segments radially increasing in size. Preliminary studies [4.1.39] have shown that this concept is qualitatively superior to a competitive concept based on a passive, thermally invariant carrier structure made of C/SiC.

The possibility of effectively damping vibrations by means of adaptive structural elements opens the future option of an **adaptive interface** between launchers and satellites, which could decisively reduce the dynamic loads acting on the latter. At present (2008) there are a lot of developments in this direction but the chances of success are still uncertain.

4.1.5.5 Spinning Devices and Flywheels

Active control of the spin in all three spatial directions is essential for the attitude control of a spacecraft (see Section 4.5).

In the case of a spin-stabilized satellite, the fast rotation around the longitudinal axis, which is simultaneously the main inertia axis, must be coped with first. Particularly in the early years of space flight, there were many satellites that received an enormous spin

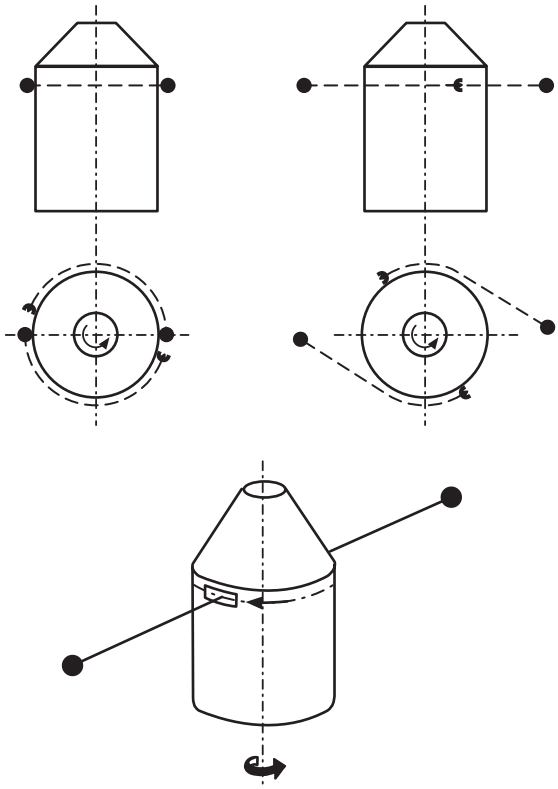


Figure 4.1.36: *Yo-Yo mechanics according to [4.1.40].*

from their launcher or from its upper stage, something which had to be eliminated for the most part to make the satellite operable in orbit. The cylindrical form of many of the early satellites (see Figures 1.1.4 and 1.1.6) facilitated the use of so-called **Yo-Yo systems**, the principle of which is depicted in Figure 4.1.36 [4.1.40].

Two small compact masses are connected by ropes to two diametrically opposed points on the cylindrical satellite structure. Initially the ropes are spooled on the structure and the masses are locked with a launch lock. As soon as this is released in orbit, the masses are centrifugally flung off as the ropes unwind. Finally the largest portion of the spin is transferred onto the distantly revolving masses, and the satellite itself rotates only slowly. The ropes are then transected with a pyroknife and the masses flap away (becoming space debris, but this was irrelevant in the early years of space flight).

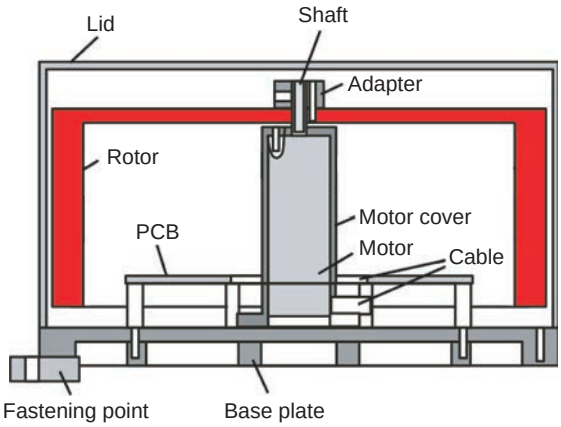


Figure 4.1.37: *Flywheel for attitude stabilization (simplified cross-section).*

Today satellites are no longer injected into orbit with unwanted high spin rates. Moreover, only some of them are still cylindrical. Primarily one relies on **flywheels** or **momentum wheels** on-board to control the spin. In low Earth orbits **magnetorquers**, which utilize the terrestrial magnetic field for attitude regulation, are also often used, but not considered here because they work purely electromagnetically and are not defined as mechanisms.

Flywheels and momentum wheels are commercially available as COTS in many forms and sizes. Figure 4.1.37 shows a typical design. The rotor is formed such that its circumferential cylindrical wall contains the largest portion of the mass in order to maximize the MoI. The inner parts are as light as possible; here any additional mass would be wasted. Such flywheels are mostly driven by BDC motors (see Section 4.1.5.4). The bearing must be very smooth running and function over long periods under space conditions (see Section 4.1.5.7 below).

The smallest available flywheels are only a few centimeters in size; one of the largest types ever built is shown in Figure 4.1.38.

4.1.5.6 Tribological Materials

Assembly parts which are moved relative to each other need to be effectively fastened and guided. The involved materials are subject to **tribological stresses** since



Figure 4.1.38: *ISS gyroscope (Source: NASA).*

they are exposed to sliding and rolling movements which occur under certain speeds and loads. **Tribosystems** are, for instance, bearings in pumps, gears, motors, as well as all kinds of hinges and connection elements which are summarized in space technology under the term **mechanisms**.

In general terms, **tribological systems** (or tribosystems) consist of assembly parts which are in contact and can be moved relative to each other (see Figure 4.1.39). The characteristics of such a system

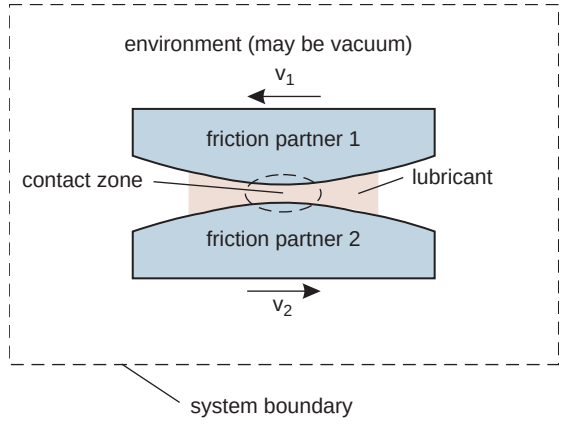


Figure. 4.1.39: *Tribological system.*

are governed by the eventual existence of lubricants and by the environmental conditions, for space applications in particular by the high vacuum. The input variables of a tribosystem are load and speed, while friction and wear are the output variables. A simplified version of the **laws of friction** is stated below:

1. The friction is proportional to the normal force. The relationship between frictional force F_R and normal force F_N defines a friction coefficient μ ($F_R = \mu \cdot F_N$).
2. The frictional force is independent of the size of the contact surface.
3. The frictional force is independent of the speed.

Friction is generated by **adhesion** and **abrasion** between the surfaces involved, and is influenced by their hardness, elasticity and roughness. The above formulated laws of friction are a simplification valid under the assumption that both the measuring periods and the scales are sufficiently large. If short-term phenomena (occurring in milli-, micro- or nanoseconds) or measurements in the micro- or nanometer range are considered, then the validity of these laws is limited. This is known as the scale dependency of the laws of friction. Deviations are also observed under vacuum conditions, where the friction coefficient may drastically increase or fluctuate, while under atmospheric conditions the surface is always covered by a protective oxide and adsorption layer, which also regulates the friction process. Tribological materials should not just have a low friction coefficient, rather also the fluctuations of this coefficient across the life span of the system should be low. Drives need to be designed for the maximal frictional resistance occurring, so that a permanently low friction coefficient makes it possible to build weight-saving drives. The independence of the friction from speed does not exist for fluids where the internal friction increases with speed. Therefore, tribosystems which are lubricated with liquid lubricants have a speed-dependent friction.

Apart from friction, **wear** is the second important parameter for the description of a tribological system. Like friction, wear can be caused by adhesion and abrasion, but also other forms, such as wear caused by fatigue, erosion or tribochemical reactions, are known. Both wear and friction are system properties which can be disproportionately influenced by the variation of other system properties. Generally, wear increases with load, is proportional to the wearing distance, and decreases with growing hardness of the surfaces involved. In addition to the material and geometrical characteristics of the friction partners, the lubricant is another important design element that influences friction and wear. For tribosystems in space technology it is mandatory to retest them after every modification of the construction or of the operational conditions, because the complexity even of simple tribosystems is still too high to accurately predict friction and wear effects under space conditions with today's calculation tools.

Tribological systems in space technology are a focal point of engineering research, because there are certain particularities compared to terrestrial tribosystems, for instance the operation in high vacuum, the influence of large temperature gradients and of radiation, and the vibrational launch loads. While the environmental pressure in the geostationary orbit lies at 10^{-13} mbar, this value may grow in the close vicinity of a satellite to approximately 10^{-10} mbar due to the outgassing of materials. Therefore space materials must be qualified for use in a vacuum; that is, they must have an outgassing rate as low as possible and be as resistant as possible against degradation by radiation or highly energetic particles (see Section 4.1.1.4 above). While metals usually satisfy these criteria, the use of polymers is more critical since monomers and additives can easily escape. This is also the case for any lubricants used. Either dry lubricants or high molecular liquids with limited vaporization should therefore be used.

When the spacecraft rotates, or when it passes through an eclipse, temperature differences of up to 100 K can easily occur. This has to be considered not only with respect to the design of mechanisms and bearings, but also for the selection of an adequate liquid lubricant, because liquids are strongly affected by large temperature gradients. Oil-based lubricants harden under low temperatures, whereas high temperatures accelerate the evaporation process, which can lead to a drying up of the lubricant. Certainly the influence of vacuum and temperature on the friction resistance can be investigated by thermal vacuum tests (TV tests), but long-term prediction of the frictional

behavior remains problematic. Therefore the trend is to control the climatic conditions for liquid lubricated systems, such as bearings and momentum wheels, by suitable sealing.

Mechanical loads in orbit are usually lower than on the ground due to the absence of gravity. Nevertheless, bearings are subject to inertia forces when the mechanism is accelerated, and optionally to prestresses which may have been applied in order to protect the assembly parts during the launch phase. Otherwise the vibrations during launch may lead to damage ("brinelling") when neighboring surfaces impact each other, an effect which is particularly critical for rolling bearings. A resolution of this issue is to apply elastic prestresses or to use damping elements.

As for terrestrial applications, the choice of the contact or bearing principle has to be made prior to the dimensioning of a space mechanism. The fundamental distinction is between sliding bearings and rolling bearings. While simple, slowly moving or seldom/single-use mechanisms are usually equipped with sliding bearings, rolling bearings are preferred for permanently moving units. Accordingly, the hinges used for the deployment of solar array panels are equipped with sliding bearings, while for instance the drives in the solar array drive mechanisms (SADM, or BAPTA, respectively, see Section. 4.1.2.2) have rolling bearings, as well as gyros for energy storage or attitude stabilization.

Suitable materials for tribologically stressed assembly parts are selected metals, polymers, composites and ceramics. Since their qualification requires considerable effort it is preferable to use already well-known and proven materials, for instance hardened steels such as AISI 52100 (100Cr6) or AISI 440C (102CrMo17). AISI 440C is a corrosion-resistant aerospace steel which, depending on its thermal treatment, covers a temperature range from −270 to +300 °C. Also, innovative nitrogen-alloyed steels such as X30CrMoN15.1 are used, which are quite resistant to wear effects in rolling bearing contacts and already proven in aerospace applications. Steels excel in their strength, ductility, hardness and ease of processing, while their disadvantage is their relatively high weight. Aluminum and titanium alloys are lighter but, due to their low hardness, suitable only for tribological purposes if the surface has a protective layer against wear. This is particularly true for titanium, which shows a broad scatter of the friction coefficient under vacuum conditions and easily tends to **cold-welding**. Tribologically used assembly parts made from titanium are either lubricated or coated.

Soft-metal alloys, like bronze or brass, are used for the cages of rolling bearings. In order to minimize wear of such soft metals a protective layer of titanium carbide (TiC), titanium nitrite (TiN), tungsten carbide (WC), diamond-like carbon (DLC) or tungsten carbide/carbon (WC/C) can be used. Such layers are applied by PVD- or CVD-coating methods. Copper alloys (bronze, brass) are also used for sliding components which can be coated with soft-metal layers (silver, lead). An overview of the tribological materials and selection strategies used in space technology can be found in [4.1.44] and [4.1.46].

4.1.5.7 Lubrication of Bearings and Mechanisms

In space there is a high risk of unintentional **welding** between metallic surfaces because, other than in an atmosphere containing oxygen, there is no protective oxide layer, and also because impacting high-energy particles support the generation of local interatomic bonds. It is accordingly important to separate the surfaces with lubricants or layering systems. Basically, one can distinguish between **dry lubricants** and **liquid lubricants**. Liquids create a lubricating film between the contact partners which regulates itself according to speed, viscosity and mechanical load. The minimum thickness of this film depends on the roughness of the surface. For metallic and ceramic surfaces the mean value of roughness is usually between 0.02 and 0.1 μm. In principle the surface should be as smooth as possible; however, it is observed that for roughnesses smaller than 0.05 μm the quality of the lubricating film can deteriorate. Lubricants where the viscosity increases under high pressure are known as shear thinning lubricants; they facilitate the separation of the surfaces even in highly stressed rolling bearings. The load in rolling bearings is transmitted across a very small surface of only a few square millimeters. The maximum compression occurring in the contact zone is known as **Hertzian stress** and reaches values up to 4000 N/mm². Rolling bearings are usually operated in a range of 2000–3000 N/mm². Under these conditions

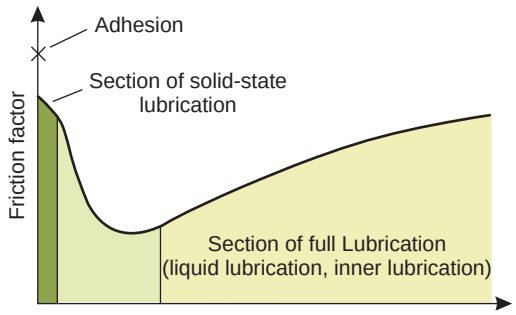


Figure 4.1.40: *Stribeck curve.*

the lubricating film must safely separate the surfaces of rolling elements and runway. Suitable algorithms and computer programs for selecting an appropriate lubricant and calculating the required viscosity are available; see [4.1.42] and [4.1.46].

The characteristic behavior of the friction coefficient as a function of the (rotational) speed under constant load and viscosity can be described by a **Stribeck curve** (Figure 4.1.40). For liquid lubrication one can distinguish between dry friction (boundary lubrication), mixed friction and the range of the completely separating lubricating film (hydrodynamic lubrication). For many friction partners the transition from resting state to movement is coupled with a sudden decrease in friction. The force necessary to overcome the resting state is known as **static friction**.

Dry lubricants are always used in the range of dry friction, where the difference between static and sliding friction must be as small as possible. Bearings operated in the range of hydrodynamic lubrication are practically free of wear. Nevertheless, due to aging processes in the lubricant itself, the life span of a fully lubricated bearing is limited; however, the estimation of a lubricant's durability is difficult. Durability tests on lubricants are time consuming and sophisticated. The possibilities for quick tests are very limited since the operating conditions strongly affect aging.

Table 4.1.4 compares the essential characteristics of dry and liquid lubricants; more detailed information can be found in [4.1.46]. Dry lubricants are used under either very high or very low temperatures. There is little risk that they could contaminate surrounding surfaces, and they can neither creep nor

Table 4.1.4: *Comparison of dry and liquid lubricants.*

Dry lubricants	Liquid lubricants
No outgassing in vacuum	Outgassing depends on vapor pressure
Large temperature range	Viscosity, evaporation depending on temperature
No creeping, no sealing necessary	Tend to creep, sealing necessary
Friction independent of speed	Friction depends on rotational speed (Stribeck curve)
Abrasion leads to fluctuations in the friction coefficient	Constant friction coefficient
Life span depends on friction wear	Life span depends on aging of lubricant

wet any surrounding assembly parts. Their estimated life span can be determined by short-term tests and depends in particular on the acting load and the wearing distance. Typical dry lubricants used to improve friction behavior are soft-metal layers of silver, lead or gold, or layers of molybdenum disulfide (MoS_2), polytetrafluorethylene (PTFE) or graphite. Layers of metal can be applied galvanically or by **physical vapor deposition** (PVD) coating [4.1.47].

Dry lubricants such as MoS_2 can also be applied by PVD. This is a sputtering process where solid material is evaporated and accelerated in the direction of the substrate. The material condenses on the surface and forms a cohesive layer. Actively lubricating MoS_2 layers have a thickness of only a few micrometers. Like graphite, MoS_2 has a lamellar structure. Under compression the MoS_2 lamellae align in parallel and glide on each other, an effect which is observed in vacuum as well as under atmospheric conditions. However, under atmospheric conditions the friction coefficient is several times larger than in vacuum, and the wear increases accordingly. Tests on MoS_2-coated assembly parts are thus carried out under inert gas atmospheres (e.g., with nitrogen). An opposite effect can be observed with dry lubricants on a graphite base where the friction coefficient decreases in the presence of humidity. MoS_2 is therefore one of the most frequently used dry lubricants in space technology.

MoS_2 can be applied both automatically and manually. A soft cloth or brush is used to distribute the dry lubricant in powder form. It can be used for both rolling bearings and sliding contacts. Particularly for rolling bearings with MoS_2-coated runways it is mandatory to check their friction coefficient by initial tests before they are finally installed. It takes some time before the MoS_2 lamellae are aligned under load and the friction reaches its minimum value, which is then maintained throughout the life span of the lubricant layer. The life span of an MoS_2 layer is limited by abrasive wear. Bearings with an MoS_2 coating are often equipped with cages which allow for dry lubrication as well (e.g., made of PTFE).

Polymers suited for dry lubrication can be utilized for sliding contacts. PTFE is frequently used and can be reinforced optionally with glass fibers; in some cases it contains an admixture of MoS_2. The low friction of PTFE-lubricated components can be explained by the transfer of a small amount of PTFE onto the opponent body (transfer lubrication). In a vacuum the friction coefficient remains practically invariant. Contrary to most other dry lubricants, PTFE has a static friction coefficient which is equal to the sliding friction coefficient. PTFE is therefore particularly suitable for moderately loaded, slowly moving mechanisms requiring low initial friction. It can be used for temperatures up to 280 °C and the compressive load should not exceed 800 N/mm². Areas of application include sliding bearings and hinges, but also rolling bearing cages because the latter are also subject to sliding strain. Apart from PTFE, polyimides are also used, in particular for highly stressed components, but their sliding properties are worse than those of PTFE. Polyamides, on the other hand, are used only rarely. Polymers are used in most cases without any additional coating; however, their load-bearing capability is significantly lower than that of metallic materials.

Current investigations are not only focused on an improved separation of the surfaces by means of dry lubricants, but also focused on material pairings based on different material classes, which helps to minimize adhesive wear. Pairings of ceramics and metals as well as of ceramics and polymers are particularly suited. Bearings composed of steel rings and ceramic rolling elements are called hybrid bearings (Figure 4.1.41).



Figure 4.1.41: *Hybrid angular ball bearing with silver coating for dry operation (Source: CEROBEAR GmbH).*

They can be operated with extremely small amounts of lubricant, and some of them can also operate in low-viscous media. Investigations on completely dry running bearings in a vacuum have proven that under certain circumstances hybrid bearings can operate even without any lubrication. In these cases the rolling elements are made of silicon nitride (Si_3N_4) or zirconium oxide (ZrO_2). In the case of the Space Shuttle, for instance, the replacement of steel bearings in the turbopumps of the main engine by hybrid bearings has increased the life span of these bearings by a factor of 60. In this application the only lubricating medium is the propellant itself; that is, the bearings are operated in cryogenic liquid hydrogen at −253 °C. Research and development work focused on hybrid bearings in space mechanisms is ongoing.

Liquid lubricants are still indispensable for applications where high reliability and durability are required and the speed is high. The design of the appropriate bearings is aimed at operation in the hydrodynamic (thick-film) lubrication range, for which the bearing size and geometry, load, rotational speed and lubrication need to be harmonized. Strong deviations from the optimal operating point lead to increased friction and wear if the bearing is operated in the mixed friction range. Lubricants for space applications should have a minimum evaporation rate in vacuum. Hence high molecular and low-volatility

lubricants are used, the most important ones being oils and greases based on perfluoropolyether (PFPE), occasionally also on polyalphaolefine (PAO). Conventional hydrocarbon lubricants are seldom used. Under certain conditions, however, PFPE lubricants tend to accelerate aging due to chemical reactions between the lubricant and the material of the friction partners. Aging is additionally accelerated by operation in the mixed friction range, and by using aluminum and titanium surfaces, low-alloyed steels, high loads and high temperatures. In contrast the use of ceramics or high-alloyed steels, limited loads and moderate temperatures contribute to a prolongation of the life span of PFPE lubricants. They have good adhesion on metallic surfaces and tend to creep. Creeping can be avoided by the construction of creep barriers (labyrinths) or by adhesive lacquers with a low surface energy. A more elaborate overview regarding the use of lubricants and their trade names can be found in [4.1.46].

Bibliography

[4.1.1] ESA/ESTEC (eds.) Aide Memoire on Structural Materials and Space Engineering. *ESA-PSS-03-212*, Issue 1, Noordwijk, 1995.

[4.1.2] ESA/ESTEC (eds.) BepiColombo Experiment Interface Document, Part A (EID-A). *SCI-PB-RS-1140 (Draft)*, Noordwijk, 2004.

[4.1.3] ESA/ESTEC (eds.) Structural Materials Handbook. *ECSS-HB-304*, 2008.

[4.1.4] ASM International (ed.) *Metals Handbook, Volume 2, Properties and Selection: Non-ferrous Alloys and Special Purpose Materials*, Tenth Edition. Metals Park, OH: ASM, 1999.

[4.1.5] Hussey, R., Wilson, J. (eds.) *Light Alloys Directory and Handbook*. London: Chapman and Hall, 1998.

[4.1.6] *Carbon and High Performance Fibres Directory*, Sixth Edition. London: Chapman and Hall, 1995.

[4.1.7] ASM International (ed.) *Engineering Materials Handbook, Volume 1, Composites*. Metals Park, OH: ASM, 1987.

[4.1.8] Ehrenstein, G.W. *Faserverbundkunststoffe*, 2nd Edition. Munich: Carl Hanser Verlag, 2006.

[4.1.9] Menges, G., Haberstroh, E., Michaeli, W. *Werkstoffkunde Kunststoffe*, 5th Edition. Munich: Carl Hanser Verlag, 2002.

[4.1.10] Heissler, H. *Verstärkte Kunststoffe in der Luft- und Raumfahrttechnik*. Stuttgart: Kohlhammer, 1986.

[4.1.11] Plantema, F.J. *Sandwich Construction*. New York: John Wiley & Sons, Inc., 1966.

[4.1.12] Allen, H.G. *Analysis and Design of Structural Sandwich Panels*. Oxford: Pergamon Press, 1969.

[4.1.13] Altenbach, H., Altenbach, J., Rikards, R. *Einführung in die Mechanik der Laminat- und Sandwichtragwerke*. Stuttgart: Deutscher Verlag für Grundstoffindustrie, 1996.

[4.1.14] Michaeli, W., Huybrechts, D., Wegener, M. *Dimensionieren von Faserverbundkunststoffen*. Munich: Carl Hanser Verlag, 1994.

[4.1.15] BEOSAT Projektunterlagen der Arbeitsgemeinschaft ERIG am Institut für Luft- und Raumfahrtsysteme der TU Braunschweig, 2006.

[4.1.16] Menges, G., Michaeli, W., Mohren, P. *How to Make Injection Molds*. Munich: Carl Hanser Verlag, 2001.

[4.1.17] Kleineberg, M., Herbeck, L., Schöppinger, C. Advanced Liquid Resin Infusion – A New Perspective for Space Structures. Proceedings, European Conference on Spacecraft Structures, Materials and Mechanical Testing, Toulouse, December 11–13, 2002.

[4.1.18] Kleineberg, M., Herbeck, L., Schöppinger, C. Industrialisierung der Prozesskette Harzinfusion. 9. Nat. SAMPE-Symposium, Clausthal, 2003.

[4.1.19] Meyer, M., Herbeck, L. Microwave Effects on CFRP Processing. 26th SAMPE Europe Conference and Exhibition, JEC, Paris, April 5–7, 2005.

[4.1.20] Marsal, O., Venet, M., Counil, J. L. *et al.* The NetLander Geophysical Network on the Surface of Mars: General Mission Description and Technical Design Status. Proceedings (IAF-01), 52nd International Astronautical Congress, Toulouse, October 1–5, 2001, and *Acta Astronaut.*, **51** (1–9), 379–386, 2001.

[4.1.21] ESA/ESTEC (eds.) Thermal Vacuum Outgassing Test for the Screening of Space Materials. *ECSS-Q-70-02A*, Noordwijk, 2000.

[4.1.22] Rabek, J.F. *Polymer Photodegradation*. London: Chapman and Hall, 1994.

[4.1.23] Gorney, D.J. *et al.* The Space Environment and Survivability. In Larson, W.J., Wertz, J.R. (eds.), *Space Mission Analysis and Design*, Second Edition. Dordrecht: Kluwer Academic and Torrance, CA: Microcosm Inc., 1992.

[4.1.24] ESA/ESTEC (eds.) Insert Design Handbook. *ECSS-HB-306*, 2008.

[4.1.25] Block, J., Brander, T., Lambert, M. *et al.* Carbon Fibre Tube Inserts – A Light Fastening Concept with High Load Carrying Capacity. Proceedings, European Conference on Spacecraft Structures, Materials and Mechanical Testing, Noordwijk, May 10–12, 2005, *ESA SP-581*.

[4.1.26] Romberg, O., Bodendieck, F., Block, J. *et al.* NetLander Thermal Control. Proceedings, 5th International Conference on Low-Cost Planetary Missions, ESTEC, Noordwijk, September 24–26, 2003.

[4.1.27] Sarafin, T.P., Larson, W.J. (eds.) *Spacecraft Structures and Mechanisms – from Concepts to Launch*. Dordrecht: Kluwer Academic and Torrance, CA: Microcosm Inc., 1995.

[4.1.28] Herbeck, L., Leipold,, M., Sickinger, C. *et al.* Development and Test of Deployable Ultra-Lightweight CFRP Booms for a Solar Sail. Proceedings, European Conference on Spacecraft Structures, Materials and Mechanical Testing, Noordwijk, November 29–December 1, 2000, *ESA SP-468*.

[4.1.29] Sickinger, C., Breitbach, E. Ultra-Lightweight Deployable Space Structures. Proceedings, 4th International Conference on Thin-Walled Structures, Loughborough, England, June 22–24, 2004.

[4.1.30] Sickinger, C., Herbeck, L., Ströhlein, T. *et al.* Lightweight Deployable Booms: Design, Manufacture, Verification, and Smart Materials Application. Proceedings (IAC-04-I.4.10), 55th International Astonautical Congress, IAF/IAA/IISL, Vancouver, Canada, October 4–8, 2004.

[4.1.31] Sickinger, C., Herbeck, L., Ströhlein, T. *et al.* Strukturmechanik entfaltbarer CFRP Booms: Analyse, Herstellung, Verifikation und Anwendung adaptiver Konzepte. Deutscher Luft- und Raumfahrtkongress (DGLR), Dresden, September 20–23, 2004.

[4.1.32] Bathe, K.J. *Finite-Elemente-Methoden* (Übersetzt von Zimmermann, P.), 2nd Edition. Berlin: Springer Verlag, 2002.

[4.1.33] Mayr, M., Thalhofer, U. *Numerische Lösungsverfahren in der Praxis*. Munich: Carl Hanser Verlag, 1993.

[4.1.34] Harris, C. (ed.) *Shock and Vibration Handbook*, Fourth Edition. New York: McGraw-Hill, 1995.

[4.1.35] Cruise, A.M., Bowles, J.A., Patrick, T.J. *et al. Principles of Space Instrument Design*. Cambridge: Cambridge University Press, 1998.

[4.1.36] Pfeil, D. NetLander Petal Opening and Uprighting Mechanisms. *Abschlussbericht Phase B*, Giesen/Hildesheim, 2002.

[4.1.37] Phinney, D.D., Britton, W.R. Developing Mechanisms. In Sarafin, T.P., Larson, W.J. (eds.), *Spacecraft Structures and Mechanisms – from Concepts to Launch*. Dordrecht: Kluwer Academic and Torrance, CA: Microcosm Inc., 1995.

[4.1.38] Wierach, P., Monner, H.P., Schönecker, A. *et al.* Application Specific Design of Adaptive Structures with Piezoceramic Patch Actuators. In McGowan, A.-N. (ed.), *Smart Structures and Materials 2002*, SPIE Proceedings, Vol. 4698, 2002.

[4.1.39] Dürr, J.K., Herold-Schmidt, U., Zaglauer, H. *et al.* Smart Composites for Adaptive Satellite Mirrors. Proceedings, Adaptronic Congress, Potsdam, April 4–5, 2000.

[4.1.40] Hallmann, W. Beispiele ausgeführter Mechanismen und konstruktiver Details. In Hallmann, W., Ley, W. (eds.), *Handbuch der Raumfahrttechnik*, 2. Edition. Munich: Carl Hanser Verlag, 1999.

[4.1.41] Czichos, H., Habig, K.-H. *Tribologie Handbuch*. Wiesbaden: Vieweg Verlag, 1992.

[4.1.42] Möller, U.J., Boor, U. *Schmierstoffe im Betrieb*. Düsseldorf: VDI Verlag, 2004.

[4.1.43] Fusaro, R. Preventing Spacecraft Failures due to Tribological Problems. *NASA/TM-2001-210806*.

[4.1.44] Jones, W., Jansen, M. Space Tribology. *NASA/TM-2000-209924*, Glenn Research Center, May 2001.

[4.1.45] Bursey, R. *et al.* Advanced Hybrid Rolling Element Bearings for the Space Shuttle Main Engine High Pressure Alternate Turbopumps. 32nd AIAA/ASME/SAE/ASEE Joint Propulsion Conference, Lake Buena Vista, FL, July 1996.

[4.1.46] AEA Technology (ed.) *Space Tribology Handbook*, Warrington: ESTL, 1997.

[4.1.47] Mertz, J. *Praxishandbuch moderne Beschichtungen*. Munich: Carl Hanser Verlag, 2001.

4.2 Electrical Power Supply

Wilfried Ley and Reinhard Röder

Electrical power is needed for the operation of all active spacecraft systems and subsystems. The **electrical power system** (EPS) of a spacecraft comprises power conversion, power conditioning, energy storage, overvoltage and overcurrent protection, and power distribution to the various users via the on-board low-voltage power distribution systems. A spacecraft's power system provides power levels ranging from a few watts up to the 50 kW currently distributed at voltages between 20 and 125 V. The completely integrated International Space Station (ISS) with 110 kW represents an exceptional power demand.

High-voltage technologies on spacecraft are increasingly used for the power supply of electrical propulsion systems.

The EPS has to maintain the energy supply of connected users during all mission phases and under all the environmental constraints encountered in space in a secure and maintenance-free manner.

An important requirement, at least for unmanned space vehicles, is that the EPS autonomously maintains the spacecraft power supply in all possible **failure conditions**, without operational intervention from the ground. A power failure causes in almost all cases the irreversible total loss of the spacecraft. Recovery of the spacecraft and a return to nominal operation

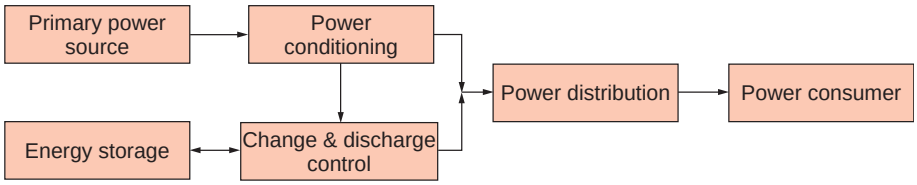


Figure 4.2.1: *The essential functional blocks of an electrical power system for a spacecraft.*

after a power loss by ground control intervention is only possible if its telecommunication system and its orbit and attitude control systems remain operational.

All these requirements call for a well-designed, robust and reliable EPS with all functions verifiable by test.

An EPS, as shown in Figure 4.2.1, can usually be broken down into four functional blocks: primary power source, energy storage, power management (power conditioning and charge and discharge control) and power distribution.

4.2.1 Energy Generation

There are various ways to supply spacecraft with electrical energy. Two basic methods have to be distinguished:

1. Energy is derived from the outside environment (so far only in the form of converted solar energy).
2. An energy source is carried on-board.

The **conversion of the primary energy** into electrical energy is performed either directly by using the **photovoltaic** effect of **solar cells** as well as chemical–electrical conversion in galvanic elements (batteries or fuel cells), or indirectly by conversion of solar or nuclear energy into thermal energy. Conversion of thermal energy into electrical energy can be achieved statically by using thermal or thermionic elements. A third method is dynamic conversion via kinetic energy using magnetohydrodynamic (MHD) generators or turboelectric thermal power machines with gas (Brayton cycle or Stirling motor) and steam turbines (Rankine cycle).

Although their efficiency is low, all nuclear energy sources flown so far use **thermoelectric converters** because of their operational reliability and simple handling.

A **solar dynamic system** converting solar energy into thermal energy represents a combination of converter and energy source. Although such systems have an efficient surface-to-power ratio (i.e., generated energy per illuminated surface unit) the development of this method has so far not been pursued because of considerable efficiency improvements in solar cells combined with their economical production costs.

4.2.2 Power Sources

Space missions usually require power sources able to generate electrical power for many orbit cycles to supply electrical loads and recharge batteries. While launch vehicles use **primary batteries** for power supply because they usually have to provide power for less than 1 hour (and at most up to 4 hours), such an approach is not suitable for operating a spacecraft over a period of weeks, months or years because the energy content of batteries alone would be insufficient. Figure 4.2.2 summarizes possible power sources.

Future lunar or planetary bases as well as long-term missions will require either extremely powerful or regenerative energy systems since their power requirements will be far beyond the capability of today's power systems like the ISS's current requirement of 110 kW.

The Sun (electromagnetic solar radiation, solar corpuscular radiation), interplanetary space (interplanetary magnetic field, cosmic radiation) and the planets (chemical sources of energy, nuclear energy, physical sources of energy with wind energy, tidal energy, geothermal energy, magnetic fields) are regarded as possible primary power sources which could be used

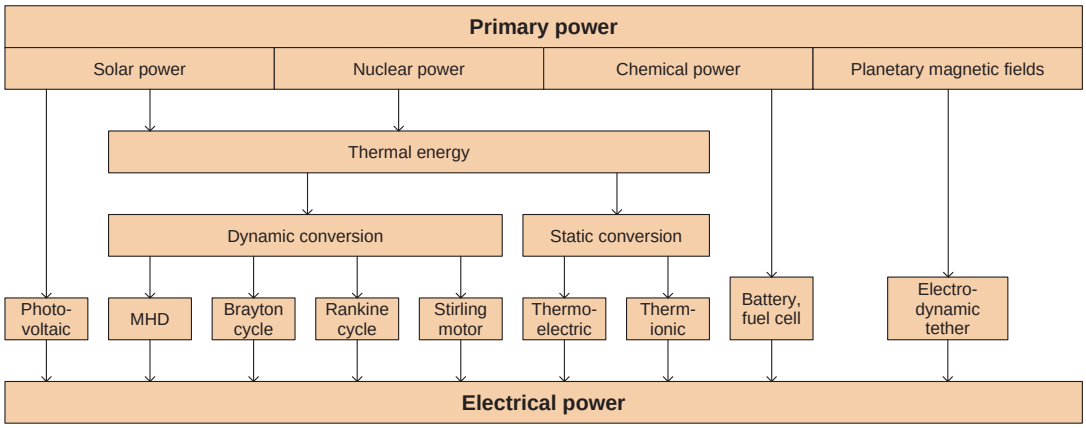


Figure 4.2.2: *Possible ways to generate electrical power.*

for the power supply of various missions, including supporting a Moon base.

However, when their physical and technical usefulness is taken into account, these alternatives are reduced to:

- Electromagnetic solar radiation
- Nuclear energy (radionuclides, nuclear fission)
- Planetary magnetic fields (electrodynamic tethers).

The power supply systems presently in use and which will also dominate the near future are based on the conversion of electromagnetic solar radiation by the following processes:

- Photovoltaic energy transformation
- Photochemical processes
- Photoelectrochemical processes
- Photobiological processes
- Thermal energy (direct utilization of thermal energy, thermodynamic processes)
- Radiation pressure.

A photovoltaic power system usually consists of the following main components:

- **Power conversion**
- **Power conditioning**
- **Power distribution**
- **Energy storage.**

It is discussed in detail in the following sections.

Up to now, the use of photochemical, photoelectrochemical and photobiological processes has not gained widespread acceptance in the field of energy processing. Thermal energy can be used either directly or via a thermodynamic cycle consisting of two energy conversion steps: thermal energy to torque, and torque to current conversion. This is described in more detail in Section 4.2.2.2.

Solar radiation pressure (solar sails, "sailing" in space) cannot be utilized for spacecraft power supply.

The processing of helium-3 cannot be considered for energy generation on the Moon (but perhaps for terrestrial fusion power plants). The same is true for helium-3 energy generation for spacecraft. Nuclear systems are used predominantly for military or interplanetary missions to serve as a possible energy source for the EPS. Detailed information follows in Section 4.2.2.3.

Tethered satellite systems with an electrically conductive tether can be used for power generation since there is an induced current caused by the movement of the tether in a magnetic field. They can also be used to produce thrust; the thrust generated by the current flowing through an electrical conductor is used to compensate aerodynamic resistance. The electrodynamic tether could be placed in a circular LEO orbit at a speed of approximately 7.6 km/s, depending on the orbital radius (Figure 4.2.3). Up to now two tether satellite missions have been undertaken by NASA in

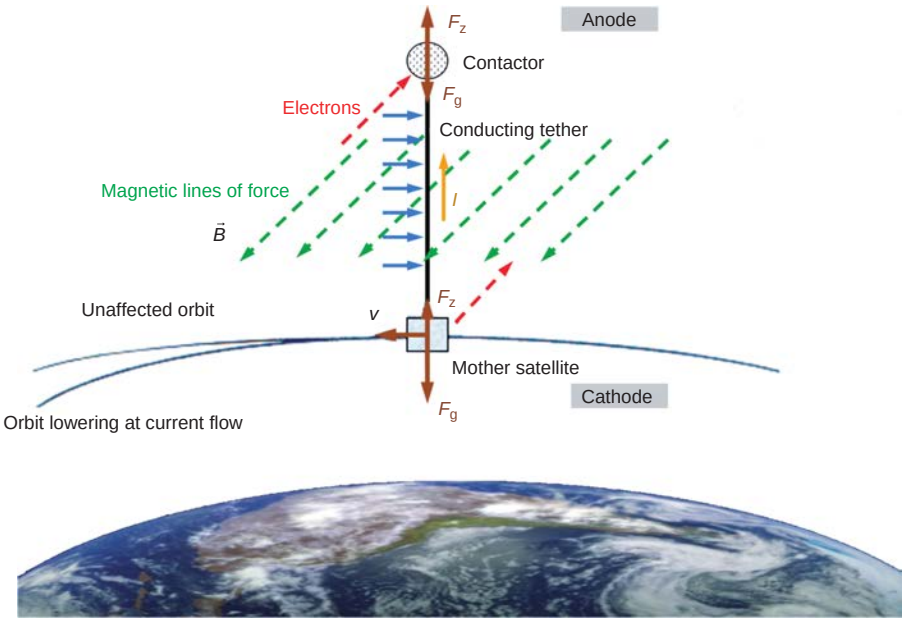


Figure 4.2.3: *Electrodynamic tether [4.2.22].*

cooperation with Italy; unfortunately they could not furnish definite proof of the physical principle because of technical failures.

Tethered applications with nonconductive tethers can be used to study aerothermodynamics (wind tunnels in orbit), for atmosphere research (temperature distribution, chemical composition, occurrences of turbulence) and, for example, to return the payload of space experiments conducted on-board the ISS.

4.2.2.1 Photovoltaics

Solar arrays (SAs) using photovoltaic assemblies (PVAs) in conjunction with rechargeable batteries (**secondary batteries**) are the most common power sources for Earth-orbiting spacecraft as well as for the ISS.

Since photovoltaics are the leading choice for providing primary power, they will be described in detail in Sections 4.2.4–4.2.7.

For the sake of completeness, the frequently made suggestion of using "solar power satellites" in geostationary orbit to convert solar energy into electrical, microwave or laser energy should be mentioned. This energy is then focused and transmitted to gigantic terrestrial antennas and reconverted to electrical energy before being fed into the local power network (5 GW). However, acceptance problems regarding the transport scenario and the financial investment have so far prohibited implementation.

4.2.2.2 Solar Dynamics

The only efficient alternative to the conversion of the Sun's electromagnetic radiation besides exploiting the photovoltaic effect is the use of solar dynamics. A solar dynamic energy supply system can use a thermal–mechanical–electrical energy converter or a conventional thermodynamic cycle process with a closed working medium cycle.

For the **solar dynamic energy conversion process** the incoming, almost parallel solar irradiation (collimation angle 32 arc minutes) is reflected by a nearly rotationally symmetric parabolic collector onto a radiation receiver whose aperture is at the focus of the parabolic collector and which is able to transfer the reflected solar energy to a working medium. This so-called processing heat is used to generate mechanical energy with the help of a thermal engine. Excess heat is radiated via thermal radiators back into space. An electric generator converts the rotational energy of the thermal engine again to electrical energy and

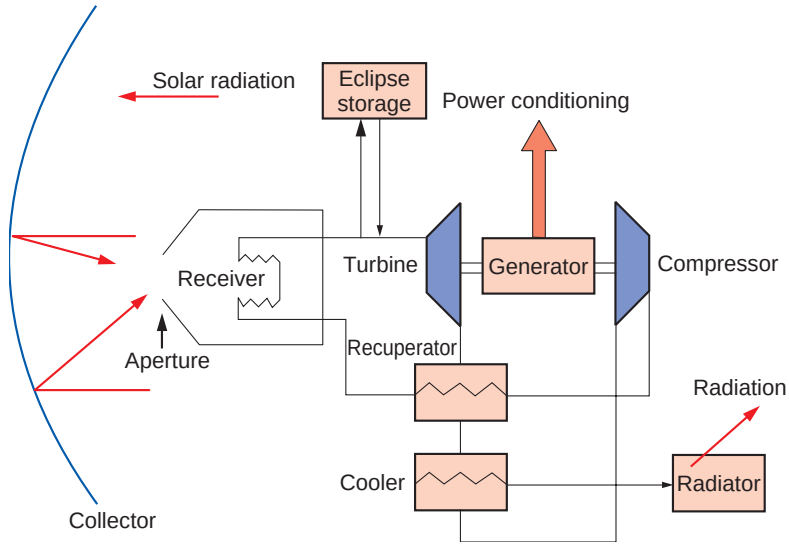


Figure 4.2.4: *Block diagram of the thermal cycle of a solar dynamic energy supply system.*

provides it to the spacecraft. A solar dynamic unit consists of typical components as shown in Figures 4.2.4 and 4.2.5.

Examples of thermal cycles useful in space applications are:

- **Stirling process** (hermetically sealed thermomechanical energy converter with external heat supply)
- **Brayton (Joule) process** (gas turbine with a monophase working medium)
- High-temperature **Rankine process** (steam turbine with biphase working medium)
- Organic Rankine process (ORC), used for medium temperature ranges.

From the T–S diagram (T = temperature in K, S = entropy) the thermal efficiency of an ideal Brayton thermal cycle with a heat exchanger can be derived as

$$\eta_{th} = \frac{q_{in} - |q_{out}|}{q_{in}} = 1 - \frac{|q_{out}|}{q_{in}} = 1 - \frac{T_5 - T_1}{T_4 - T_2} \qquad (4.2.1)$$

Using the thermodynamic relationship for an ideal gas in isentropic compression and expansion,

$$\frac{T_1}{T_2} = \frac{T_5}{T_4} = \left(\frac{P_1}{P_2}\right)^{\frac{\kappa-1}{\kappa}} \qquad (4.2.2)$$

the thermal efficiency can be formulated as

$$\eta_{th} = 1 - \frac{T_1}{T_2} = \left(\frac{P_1}{P_2}\right)^{\frac{\kappa-1}{\kappa}} \qquad (4.2.3)$$

The closed gas turbine process works as a regenerative Brayton process with an He/Xe gas mixture. The thermomechanical efficiency can be improved considerably by increasing the upper process temperature.

Concentrators with very high optical quality and concentration capabilities above 1000 are desired as **collectors**.

Because of the very high energy requirements for the first operational phase of the ISS (110 kW), for several months at the end of the 1980s NASA considered using solar dynamic energy supply systems to comply with a requirement to minimize collector surface area. There was concern about the additional fuel needed to compensate the aerodynamic drag; for example, for accelerating the space station to maintain its orbit.

Solar dynamic EPS systems with an energy output per surface unit which is two to three times higher then solar generators are being further developed to achieve higher cycle robustness, reliability and longer lifetimes. They can then be used in high-performance spacecraft in LEOs of 200–400 km, because less collector area causes fewer disturbances from aerodynamic

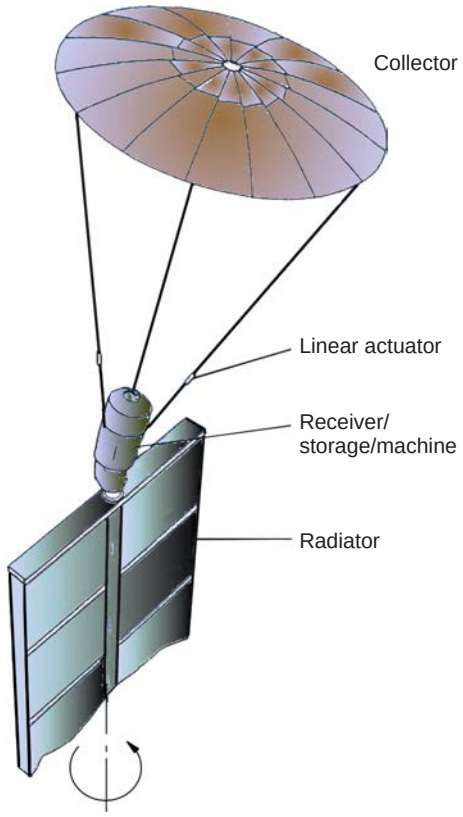


Figure 4.2.5: *Schematic of a solar dynamic energy supply unit.*

drag. On the other hand, because of their rotating aggregates, solar dynamic EPSs interfere with the microgravitation level.

The primary energy storage unit converts the energy input modulated by day–night cycles to a constant heat flow input for the thermal–mechanical–electrical converter. The latent heat accumulator, its capacity determined by the melting enthalpy of the storage medium, is considered to be the future solution because of its advantages over "sensitive" storage systems which use the high specific heat capacity of certain substances and compounds.

Another storage system is **flywheel** storage, which consists of a rotating solid body whose specific energy contents (Wh/kg) increases with the square of the angular velocity. Considerable regulation effort is, however, necessary to maintain a constant voltage supply. In addition, the rotational speed is restricted due to high centrifugal forces.

4.2.2.3 Nuclear Power Supply

The static and dynamic conversion of thermal energy generated by nuclear processes is mainly found in satellites used for military reconnaissance and surveillance missions flying in LEO, or for interstellar missions (unmanned deep-space missions). For such missions, photovoltaic systems would lead to unfavorable spacecraft configurations due to the large demand on PVA area as required to compensate for the low solar intensity resulting from the long distance to the Sun. Another application reason may be given by missions with high space radiation exposure sufficient to damage photovoltaic components.

Nuclear power supply systems include **radioisotope thermoelectric generators** (RTGs) and nuclear reactors. The latter can be further subdivided into static systems with direct energy conversion and dynamic systems using thermodynamic processes (Figure 4.2.6).

RTGs using SNAP (System for Nuclear Auxiliary Power) with direct energy conversion have an efficiency of η = 5–10% and low power outputs (P_{el} = 1 kW). These systems were used for the Galileo and Ulysses satellite missions along with a GPHS (General Purpose Heat Source) expected to be able to survive catastrophic crashes.

RTGs using dynamic isotope power systems (DIPS) are used in Brayton or ORC processes, yielding an efficiency of η = 18–25% at an output power level of 1–10 kW_{el}.

Nuclear reactors with **direct energy conversion** use thermodynamic, thermionic and AMTECs (Alkali Metal Thermoelectric Converters) with an efficiency of η = 10–19%. Dynamic energy converters yield efficiencies of η = 18–25% at power outputs of P_{el} > 20 kW.

The American test reactor model SNAP 10-A (1965, operating for 43 days in orbit) as well as the Russian RORSAT satellites use **thermoelectric energy converters**, while the Russian TOPAZ satellites use **thermionic energy converters**, similar to the not yet completely developed SP-100 reactor which provides a power output of 100 kW_{el}. The nuclear reactors use

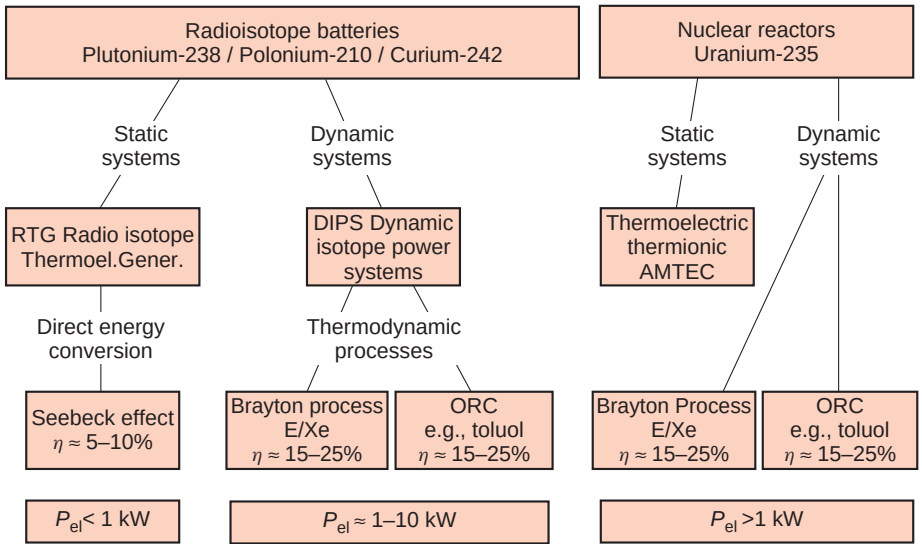


Figure 4.2.6: *Overview of nuclear energy supply systems.*

highly enriched uranium-235 or uranium oxide as fuel with a half-life of $7.1 \cdot 10^8$ years.

In 1995, with the support of DARA, the German Space Agency at that time, a German–Russian research team conducted a study called "Advanced Interplanetary Missions Using Nuclear Electric Propulsion" (Topas-25 reactor) as part of the CONNEP project (Contributions to Nuclear Electrical Propulsion for Advanced Space Missions project). In the succeeding study of CONSEP (Contributions to Solar Electric Propulsion for Advanced Space Missions) the alternatives of nuclear energy propulsion versus solar energy propulsion are currently being compared with regard to spacecraft configuration as well as to mission and orbital aspects. The study is based on the advanced German RIT (**Radiofrequency Ion Thruster**) technology.

The advantages of nuclear power supply systems are as follows:

- Continuous power provision without the demand of additional energy storage capability for eclipse power supply.
- High reliability.
- External energy sources not required.
- Resistance to high particle fluence, for example within the Van Allen Belt or in orbit around Jupiter.
- Compact design with small radiators.
- Long service life.

These have to be traded against severe **acceptance problems** like:

- Disposal of radioactive material in LEO missions (final disposal, reentry).
- Additional security measures for safeguarding the launch and powered flight.
- Shielding of the radioactive radiation.
- Reentry of a nuclear reactor into the Earth's atmosphere.

4.2.2.4 Chemical Energy/Fuel Cells

Although rarely used, and if so, mostly in human space missions, the **fuel cell** is worthy of serious consideration as an energy source alternative to secondary batteries in future unmanned missions. If it could be qualified for a long spacecraft mission life it would offer weight and cost savings because of its high energy density, achieving presently about 500 watt hours per kilogram (Wh/kg).

A fuel cell uses the classic H_2O electrolysis process in which hydrogen and oxygen produced from water is reversed, with the help of an electric current.

Figure 4.2.7 illustrates the principle of a fuel cell: two electrodes provided with oxygen and hydrogen are placed in an electrolyte (e.g., phosphoric acid) which connects them. The electrons flowing through the external circuit provide the desired electrical energy,

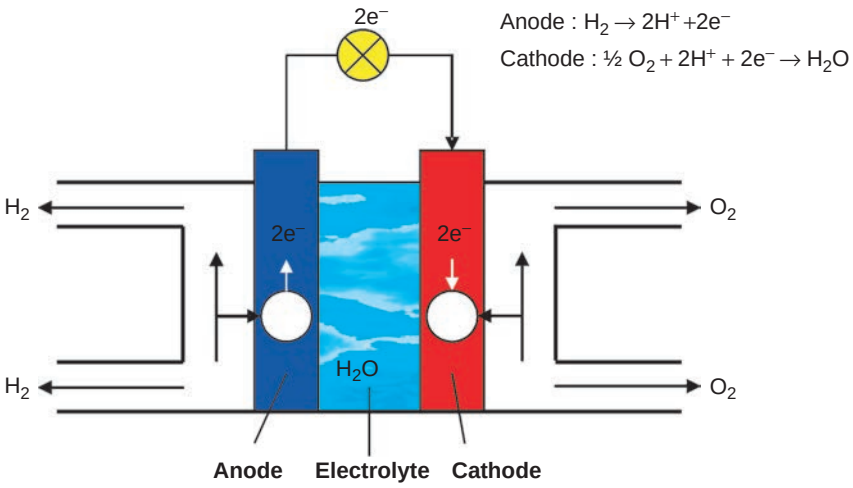


Figure 4.2.7: *Basic design of a fuel cell using an acid electrolyte.*

with water being the waste product, while the transport of charges through the electrolyte within the cell is described by the reaction equation

$$H_2 + {}^1/_2 O_2 \rightarrow H_2O$$

The process is equivalent to that in standard acid batteries or accumulators: hydrogen and oxygen in gaseous form must be provided continuously and the reaction products as well as the heat generated by the reaction must be removed from the electrolyte and the cell.

The various fuel cells can be classified by their operating temperatures (low-, medium- and high-temperature fuel cells) or by their electrolyte (water-soluble, alkaline and acid electrolytes, as well as polymer electrolyte membranes and systems with carbonate melts or ceramic oxides as electrolytes).

The energy supply system for the Space Shuttle orbiters consists of three fuel cells operating autonomously (Table 4.2.1). They represent a notable improvement over the hydrogen–oxygen battery system of Bacon (Apollo batteries) since they are 20 kg lighter in weight and their efficiency is higher by a factor of 6–8. Two cell blocks electrically connected in parallel and mechanically forming one unit, each block containing 32 single cells connected in series, yield one 28 V fuel cell battery.

4.2.3 Designing an Optimized Electrical Power System Architecture

In order to implement a robust, reliable and especially application-oriented EPS architecture for a planned space mission, and having to choose from a rather limited number of available options, a certain order in the selection process workflow is necessary (see Figure 4.2.8).

A thorough analysis of the EPS requirements together with an optimized spacecraft design not only leads to a customized design of the power sources, but also results in weight and system cost reductions. This is demonstrated very impressively by the current valid cost figures below, which are of course subject to change:

- €35–150 per gram for unmanned spacecraft
- €300–1100 per watt of installed BOL (Begin Of Life) solar array power output
- €50–350 per watt hour BOL battery energy
- €30 000–160 000 per module for power conditioning and power distribution electronics.

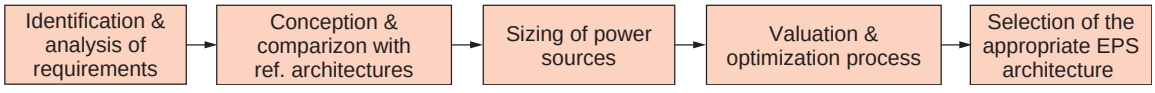


Figure 4.2.8: *Workflow for the EPS selection process.*

Table 4.2.1: *Technical data of the Space Shuttle orbiter fuel cells.*

Open-circuit voltage for one cell	≈ 1.2 V
Voltage at nominal load	0.9 V
Nominal load (100 mA/cm^2)	40.0 A
Electrode surface	400 cm^2
Electrode diameter	225 mm
Electrode density	2.0–2.5 mm
Working gases	H_2 (liquid), O_2 (liquid)
Operating pressure of the gases	40–60 bar
Operating temperature	200–230 °C
Cells per battery block	31
Nominal voltage	28 V
Nominal output power	1.12 kW
Block weight The fuel cell arrangement consists of three units of 110 kg each and two gas tanks of 240 kg each, in total 810 kg	110 kg
Flight duration	10 days
Electrical energy used during the flight	500 kWh
Energy density	617.3 Wh/kg
Power density for the flight duration	4.15 W/kg
Permanent power output of one cell block	2.3 kW
Peak power output of one cell block	6 kW
Nominal load of one cell block	82 A
Permanent power output of the total cell arrangement	14 kW
Peak power output of the total cell arrangement	36 kW
Electrical energy required for a three-day mission	1000 kWh
Fuel consumption during a three-day mission:	
• Gaseous hydrogen (cryogenic, liquid)	≈ 480 kg
• Gaseous oxygen (cryogenic, liquid)	≈ 480 kg
• Generated H2O (reaction product)	≈ 600 liters

4.2.3.1 Identification of the EPS Design Requirements

The EPS design requirements are basically determined by the following:

- The spacecraft's **power requirements**, which have to be covered in all mission-dependent operational modes and during spacecraft anomalies, also taking into account the degradation of the power sources over the mission lifetime.
- **Orbital parameters** like altitude, eccentricity, inclination, local time (defined by the ascending and descending node track when crossing the equator) and the resulting durations of sunlit and **eclipse** phases per orbit. Solar power systems must be designed such that in sunlight both the power users are sufficiently supplied and enough energy is generated and stored for power provision during eclipses and all other phases of deficient solar illumination.
- The optimum utilization of available **solar array power and battery energy** whenever required by the power demand of the spacecraft.
- The user's power requirements regarding the **quality of the distributed power**, such as power bus service voltage range, power bus stability at load change, source impedance and interference voltages.

In addition, the EPS must be designed to provide full functionality in the case of a failure or, in human spacecraft, under double failure conditions within a single component, circuit or any other single element.

Often solar array and battery technology as well as the EPS topology are predefined by the applicable specifications of requirements.

Other important requirements are as follows:

- Optimization of the EPS design with the goal to achieve low recurring **cost for flight components** or to use off-the-shelf (OTS) components wherever possible.
- Utilization of components proven to be suitable for space flight because of their use in other successfully flown spacecraft, in order to lower insurance cost.
- The highest possible numerical and functional reliability of the EPS in order to comply with up to 15 years operational life in orbit, by applying appropriate **redundancy concepts** which use unit and circuit redundancy to a necessary, but not overdesigned extent.
- Fulfillment of all functional and human-safety-related requirements.

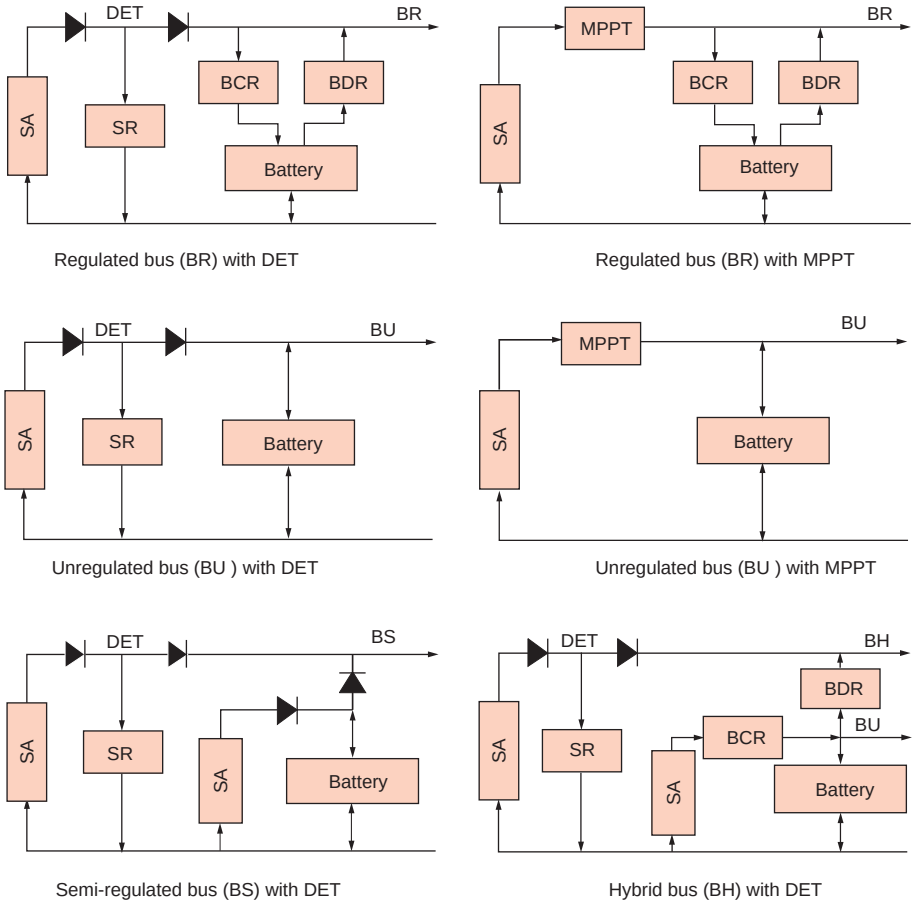


Figure 4.2.9: *Most common EPS architectures for spacecraft (SA = Solar Array).*

- Facilitation of an architecture with growth potential allowing a moderate increase of spacecraft power demand.
- Implementation of a design which can be completely tested under flight-representative conditions.
- Increased hardness against space radiation.

All these requirements must be carefully analyzed and evaluated. Should they not be detailed enough to allow the EPS engineer to design the required EPS architecture and to size the electrical power sources, missing design requirements and parameters have to be established by referring to EPS design standards and space technology handbooks, as well as by comparisons to reference architectures.

4.2.4 Electrical Power System Architectures

The required concept for **power management and distribution** (PMAD) is essentially determined by the EPS architecture. State-of-the-art PMAD concepts use either regulated (constant) or nonregulated (varying) primary power buses, or, in a combination of both methods, a semiregulated or hybrid power supply system. The primary power of the spacecraft, in most cases generated by a (photovoltaic) solar array, will be fed into the main power bus (**main bus**) of the EPS, by either direct or indirect energy transfer, as necessary to cover the actual bus power demand.

Power transfer from the solar array (SA) to the main bus by **direct energy transfer (DET)** corresponding

to the exact amount of power necessary to satisfy the bus power demand will be achieved by regulation methods like (see Figure 4.2.9):

- **Shunt regulation** (excess SA power will be bled off by shorting the corresponding amount of photovoltaic circuits (strings))
- **Series regulation** (the SA power will be fed to the bus via a linear power regulator)
- **String switching** (strings will be switched onto the bus corresponding to the amount of power as demanded by the power users).

In the case of indirect energy transfer the regulated power is provided to the main bus by DC/DC converters. These converters are controlled as a function of the required bus power demand by extracting the appropriate operating point on the SA power characteristic up to the power maximum. Therefore, this control principle is also called **maximum power point tracking (MPPT).**

Main Bus Concepts

In general DC voltage main buses are designed to supply on-board users with power, and only in exceptional cases is AC voltage used.

Regulated Bus

The classic, completely regulated primary power bus (BR) requires an efficient control loop for the three domains of SA power conversion, battery charge and battery discharge in order to provide in each power control mode a permanently stabilized bus voltage of typically:

- 28 V at a distributed bus power up to 2.5 kW
- 50 V at a distributed bus power up to 8.0 kW
- 100 and 125 V at a distributed bus power above 8 kW.

It is recommended that bus voltages should not be selected below 20 Vdc and not exceed 125 Vdc. Bus voltages below 20 Vdc might lead to unwieldy current densities, while bus voltages above 125 Vdc might easily generate dangerous potential gradients leading to ionization of air molecules (plasma) causing coronal and electric arc discharges, in particular during tests in partially pressurized environments.

It also has to be considered that the biggest selection of OTS components is available for operating voltages between 28 and 50 Vdc.

More restrictive standards like the ESA documents PSS-02-10 and ECSS-E-20A recommend selecting the next highest available bus voltage above the value

$$\sqrt{P(V)} \qquad \text{for LEO} \tag{4.2.4}$$

and

$$\sqrt{0.5\,P(V)} \qquad \text{for GEO} \tag{4.2.5}$$

(P = primary bus power, V = bus voltage).

The advantages of the regulated bus are as follows:

- It is excellently suited to EPS systems using two or more batteries, in particular if direct parallel connection of the batteries is not allowed, such as for nickel–cadmium (NiCd) and nickel–hydrogen (NiH_2) technologies. The reason is that, with the BR concept, each single battery is operated separately by using a dedicated battery charge regulator (BCR) and battery discharge regulator (BDR), therefore the number of supply buses is independent of the number of batteries.
- Potential linear increase of the bus voltage is possible by later integration of additional electronic modules to convert SA and battery energy.
- Potentially a 1–4% higher efficiency of regulated bus-supplied user power supplies (DC/DC converters) is offered compared to an unregulated bus voltage since the converters require less input filtering effort and also can be run at their optimal operating point.
- Direct power connection to users lacking their own power supply is always possible.
- No DET-typical **power lockup**, which means the retention or binding of SA power by clamping the bus voltage to the battery voltage level.
- Low and controlled source impedance up to 100 kHz in all power bus operating modes.
- Simplified thermal design of the spacecraft (electric heater circuits deliver constant heat at a fixed bus voltage).
- Good EMC performance due to load regulation and bus impedance performance.

- It is not necessary to analyze bus voltage variations, in particular to determine the minimum bus voltage. However, the minimum battery voltage must in any case be analyzed for its compatibility with the BDR input voltage range.

The disadvantages of the regulated bus are as follows:

- Higher EPS weight and higher cost because of the additional battery charge and discharge regulators.
- Higher control loop effort for bus voltage regulation.
- More SA power required because of
  - approx. 3–13% power loss in the BDR (Battery Discharge Regulator)
  - approx. 3–10% power loss in BCR.
- The BDR must be able to cover all peak power demands on the bus during the battery discharge cycle.
- Application of passive fuses with high rated current is delicate, since the rupture current capability must be provided either by the main bus capacitance (unlikely), or by the BDR.

Application Evaluation

In many cases direct supply of primary power users from a regulated bus is not needed, since almost all bus power supplied OTS units are equipped with their own input power DC/DC converter for accommodating a wide range of input voltages for multiple applications.

If no other advantages of the BR outweigh the disadvantages, there are no cost savings in using the BR.

Standard Quality Requirements for the Regulated Bus

- **Bus voltage regulation:** ±0.5% of the nominal voltage in steady-state mode.
- **Voltage variation at load changes:** ±1% at load changes of up to 50% of the nominal load and a maximum of ±5% of the nominal voltage at all transient induced voltage swings.
- In the case of a **fuse blowing,** recovery from the fuse clearance shall not produce an overshoot of more than 10% above the nominal bus value.
- Induced voltage ripple: < 0.5% peak to peak of the nominal bus voltage.

What Is Power Lockup?

Power lockup for the example of a DET-regulated SA is an operating condition in which the SA is in fact capable of satisfying the bus power demands. It is clamped to a lower power operating point dominated by a lower voltage condition of the battery (e.g., after battery discharge at the end of an orbital eclipse), which is directly connected to the main bus. If the momentary operating point power of the SA is below the power demand on the bus, the bus remains trapped on the battery with the battery providing the difference in power between what the bus load requires and what the array can provide at the clamped voltage. The battery will be discharged further, thus reducing the bus voltage even more. Without measures to relieve the bus, the battery would be completely discharged, which might finally lead to the loss of all functions on-board the spacecraft followed by a premature termination of the whole mission.

Power lockup is possible in any EPS topology if a relative high ohmic current source like SA is directly connected to a low ohmic voltage source (battery), as is the case in unregulated bus designs in combination with a DET-regulated SA.

Power lockup does not occur with MPPT regulation.

Unregulated Bus

The concept of an unregulated bus (BU) is mainly used to simplify the on-board power supply system; however, it implies the requirement that the connected users tolerate bus voltage variations of up to ±20%.

Battery voltage dominated, unregulated bus voltages are inherently dependent on type and number of battery cells connected in series. Typical, but not standardized ranges are:

- 28 V ± 6 V at a bus power up to 2 kW
- 35 V ± 7 V at a bus power up to 3.5 kW
- 42 V ± 8 V at a bus power up to 5 kW.

Higher voltage ranges are not applied for typical satellites. However, the technique of the unregulated bus is adaptable to 125 V. Those voltage levels are needed for specific satellite subsystems like an electric propulsion unit, for example, which needs voltage levels around 100 V.

Advantages of the unregulated bus are as follows:

- EPS weight and cost savings, since no additional BCR and BDR devices are necessary. This is particularly advantageous for power supply bus operations with short but high peak loads or with impulse power profiles occurring during radar instrument operations. This advantage is significant for power bus operation at high power pulse loads, or high peak power load profiles. The BDR of a regulated bus system would have to be designed for those peak loads, while the BU can deliver any peak load above the available primary power directly from the battery without any additional discharge control.
- Less SA power is required since power losses caused by BCR and BDR do not occur.
- Higher numerical and functional EPS reliability because of reduced electronic circuitry.

The disadvantages of the unregulated bus are as follows:

- Almost all bus power users need their own **auxiliary power supply** (but this is standard for OTS equipment).
- The user's internal auxiliary power supplies are approximated 1–4% less efficient if supplied from a BU compared to BR-provided power because it does not operate at an optimized regulator duty cycle, and requires higher input filtering effort.
- Higher design effort for the main bus filter in order to obtain an EMC qualified power bus.
- The actually extracted SA power is determined by the momentary battery voltage unless additional measures are foreseen (see the description of the power lockup condition).
- Up to 50% higher losses within the power harness, or up to 25% harness weight increase if cables with larger cross sections are used to reduce power losses.
- Not suitable for EPS concepts using two or more NiCd or NiH_2 batteries.
- An analysis of the bus voltage variations is necessary to determine the minimum bus voltage.

Application Evaluation

The BU is the most preferable solution for LEO applications, where the addition of power regulators for battery charging has a significant weight and cost impact, because the required battery charge power has a level close to the total distributed bus power.

The BU is excellently suited for spacecraft with pulsed loads as generated by a **synthetic aperture radar** (SAR), or with high peak power profiles, as generated by probes, or by satellites with specific communication traffic above a congested area.

The BU is also a well-suited solution for spacecraft applying AC power supplies.

Semiregulated Bus

The semiregulated bus (BS) is adopted as a compromise between the two extremes of BR and BU, as described above, and is usually based on the desire to eliminate the weight and cost impact of the BDR on the one hand, and to eliminate as far as possible the power lockup conditions of DET-regulated SA power, on the other hand.

The BS system provides regulated bus voltage during periods of sufficient SA power (sunlight period), or else battery-dominated unregulated bus voltage during periods in which the generated SA power is not sufficient to provide the required bus power; that is, in cases where the battery is used to augment the required power (such as during eclipse phases, during peak loads temporarily exceeding SA power provision or during loss of Sun orientation of the SA).

The advantages of the semi-regulated bus are as follows:

- Provision of a BR-type stabilized main bus connected to a solar generator performing at an **optimum operating point** as soon as the generated SA power exceeds the bus power demand. In this case the SA operating point as well as the usable SA power are no longer dominated by the actual battery voltage (no risk of a power lockup condition).
- EPS weight and cost savings because no additional BDR devices are necessary.
- The BS concept could be adapted to EPS designs with two or more NiCd or NiH_2 batteries.

The disadvantages of the semi-regulated bus are as follows:

- All bus power users requiring a stabilized supply voltage, but are not able to run during periods

of sufficient SA power, need their own auxiliary power supply.
- The DC/DC converters work with less efficiency during unregulated bus periods.
- Additional design effort for the main bus filter due to EMC considerations.
- Higher control loop effort for bus voltage regulation.
- Additional circuitry required for merging of the battery charge sections of the SA with the regulated main bus, in order to allow the battery charge sections to support the main bus power provision, for example, during a transition from eclipse to sunlight. This support requires corresponding decoupling devices (e.g., diodes), which dissipate additional power and thus reduce the efficiency of the bus power regulator.

Application Evaluation
The BS is well suited for geostationary spacecraft requiring little power during eclipse operations.

The advantages of the unregulated bus are reduced if the spacecraft requires power for full payload operation also during eclipse phases.

Hybrid Bus
The hybrid bus concept was designed to fill the gap between the BR and BU approaches. Part of the SA is used to supply power to a permanently regulated bus via a power regulator. This power regulator serves two purposes: primary power conversion and transfer as well as battery discharge (see also the description of the BR functions).

The second SA section supplies its power to a BU via a BCR.

A typical hybrid bus application would be useful for a spacecraft having a combination of pulsed and static loads on-board, where the static loads use the regulated bus, while the pulsed loads are directly supplied by the unregulated (battery dominated) bus.

Advantages of the Hybrid Bus
Impulse loads, or any other electromagnetically disturbing dynamic bus loads with high power demands, and "quiet" static loads are supported by separate bus systems. Separation of the different bus load types simplifies the measures for electromagnetic cleanliness control inside the on-board power supply system.

Disadvantages of the Hybrid Bus
As in the case of the BR bus, three different controllers are needed, in particular the DET or MPPT regulator, the BCR and the BDR. The splitting of the SA power into two sections must be done very carefully, since the two buses cannot buffer each other with the available excess power without additional circuit design effort.

Application Evaluation
The hybrid bus could easily be replaced by a simpler EPS architecture which provides an unregulated bus for the dynamic loads, and supplies the static loads with the desired permanently regulated bus voltages by an additional DC/DC converter, which is directly connected to the BU.

4.2.5 Solar Array

4.2.5.1 Solar Cell Technologies

As an introduction, the basic principles of the so-called **photovoltaic effect** will be described. Essentially two processes are responsible for the conversion of sunlight into electrical energy within a photovoltaic device (solar cell): first, the absorption of solar (electromagnetic) radiation within a light absorbing semiconductor and the associated generation of charge carriers (electron–hole pairs); and second, the separation of those electrons and holes under the influence of the electric field across a semiconductor junction, thus generating an electromotive force (EMF) and a photocurrent.

Characteristic of such a **solar cell** is a large-area semiconductor with an integrated p/n-junction beneath its surface. Illumination with photons generates electron–hole pairs above and below the p/n-junction. The minority carriers, namely the electrons in the p-region and the deficit electrons (holes) in the n-region doped into the p/n-junction by diffusion, are transmitted

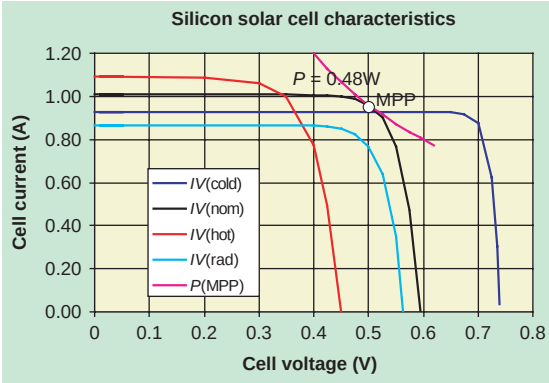


Figure 4.2.10: *Typical I–V curve of a silicon solar cell.*

Table 4.2.2: *Remaining factors R of a silicon cell for a particle fluence (electrons per cm²) of 1 MeV.*

Particle fluence	BOL	3.0E+14	1.0E+15	3.0E+15
Parameter	Nominal	*R*	*R*	*R*
V_{oc} [mV]	628	0.914	0.888	0.851
I_{sc} [mA/cm^2]	45.8	0.882	0.846	0.758
V_{mp} [mV]	528	0.912	0.885	0.844
I_{mp} [mA/cm^2]	43.4	0.876	0.837	0.757

across the p/n-junction, forced by the electric field. This causes the n-region to be charged negatively and the p-region positively, resulting in a photoelectric current through an external electric circuit.

A solar cell is characterized by the following parameters:

- Short circuit current I_{sc} (voltage $V = 0$; load resistance $R = 0$)
- Open circuit voltage V_{oc} (current $I = 0$ at infinite load resistance $R = \infty$)
- Maximum power point current I_{mp} (current at maximum solar cell output power)
- Maximum power point voltage V_{mp} (voltage at maximum solar cell output power)
- Fill factor $(I_{sc} \times V_{oc})/(I_{mp} \times V_{mp})$.

Further important properties result from changes in the above parameters caused by temperature and cosmic radiation (electrons, protons and alpha particles).

Figure 4.2.10 shows the typical **current–voltage characteristic** (*I–V* characteristic) of solar cells, in this case for a standard silicon cell with an active surface area of 26 cm^2 and a conversion efficiency of 13.5 %. The **maximum power point** (MPP) is at 0.48 W. The fact that this point is indeed the MPP is demonstrated by the tangent line, representing the constant value of 0.48 W, which intersects the nominal *I–V* curve only in the MPP. Short circuit current I_{sc} and the maximum power point current I_{mp} increase proportionally to the photon intensity. Both currents increase insignificantly but linearly with temperature. The open circuit voltage V_{oc} and the maximum power point voltage V_{mp} increase logarithmically with increasing photon intensity, and drop significantly, but linearly with rising temperature (compare the characteristics for IV_{cold} and IV_{hot}, each describing the variation of the *I–V* curve over a temperature change of plus and minus 68 °C related to IV_{nom}).

The MPP power increases almost linearly with the photon intensity and decreases significantly with increasing temperature. Cosmic radiation reduces the current and voltage values, depending on particle fluence and particle energy hitting the active cell material. The **degradation factors** (remaining factors, R), usually defined by the cell manufacturer for each cell type (Table 4.2.2), are related to different particle fluences, normalized to a particle energy density (electrons per cm^2) of 1 MeV.

Although monocrystalline single-junction silicone (SJ-Si) and GaAs/Ge-based multijunction gallium arsenide (MJ-GaAs) solar cells exist, basically only two types of solar cells are used in space SAs. In a few cases of recurrent SA manufacturing, single-junction and dual-junction GaAs cells are used. For the sake of completeness, gallium indium phosphide (GaInP) cells, thin-film cells of amorphous silicon (a-Si), cadmium telluride (CdTe) and copper indium gallium diselenide (CIGS) also must be mentioned. CIGS cells, which are still under development, have efficiencies of 20% measured under laboratory conditions. They are characterized by radiation insensitivity, low weight and low cost, and are very promising for future applications.

Space qualified production lines of MJ-GaAs cells currently provide 3J-GaAs cells (with three stacked

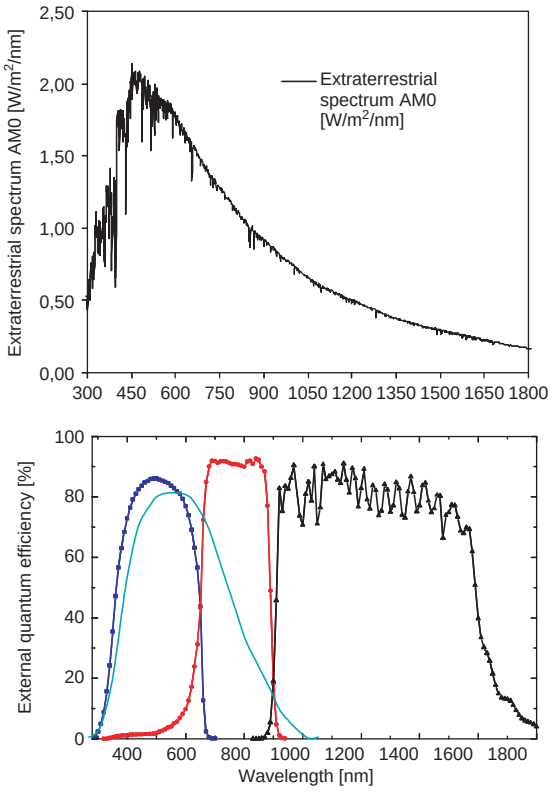


Figure 4.2.11: *AM0 solar spectrum, compared to the staggered spectral sensitivity (EQE = External Quantum Efficiency) for a TJ-GaAs cell (blue, red and black curve). The green curve shows the typical spectral sensitivity of a silicon cell (Source: [4.2.7]).*

subcells); they are also known as TJ-GaAs (triple-junction) cells. Also, 4J-GaAs und 5J-GaAs solar cell prototypes have already been successfully tested under laboratory conditions.

Whereas silicon solar cells convert the visible solar spectrum into electrical power, MJ-GaAs solar cells are also sensitive to the ultraviolet (UV) and infrared (IR) bands of the solar spectrum, as shown in Figure 4.2.11.

The main difference between both cell types is the much higher voltage range of the MJ-GaAs cells (4.5 times higher with TJ-GaAs), while its produced current is lower (approx. 0.4 times lower with TJ-GaAs). The lowest photocurrent generated by one of the three subcells of a TJ-GaAs cell determines the total current provided by the cell.

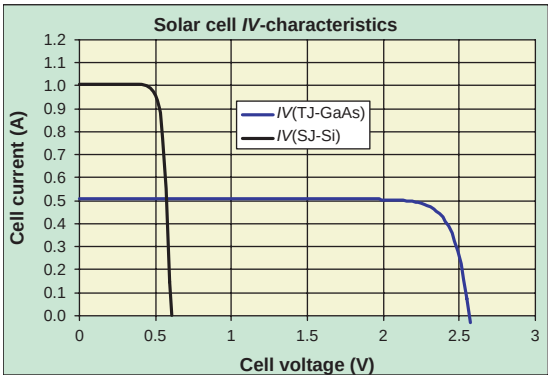


Figure 4.2.12: *I–V characteristic comparison between a TJ-GaAs and an SJ-Si cell.*

The typical efficiency of a TJ-GaAs cell of 28% must be compared to the maximum possible 17% of a so-called **Hi-Eta-Si cell**. Using 4J- and 5J-GaAs laboratory cells, actual efficiencies of 30–35% can be achieved.

Because of the much better performance of the MJ-GaAs solar cells (see also the comparison in Figure 4.2.12) one can expect that within a few years the use of silicon solar cells will be discontinued. Today new satellite designs use MJ-GaAs cells almost exclusively. The superiority of a MJ-GaAs cell is underlined by its reduced sensitivity to cosmic radiation. Typical end-of-life (EOL) efficiencies of solar cells after 15 years in geostationary orbit are (at 28 °C reference temperature)

- Hi-Eta-Si: 12.0% (30% degradation)
- TJ-GaAs: 24.5% (12.5% degradation).

4.2.5.2 Silicon Solar Cells

The **standard Si cells**, usually of 0.18 mm thickness, are built on a p-doped Si base material and a shallow p/n-junction. The BOL efficiency varies between 12 and 14%.

Further developments yielded the 0.1 mm thick Hi-Eta-Si cell with a texturized surface, allowing absorption of more sunlight through the enlarged surface area. The surface enlargement is achieved by etching pyramid-like structures of approximately 0.02 mm depth into the cell surface. Figure 4.2.13 shows the construction principle with a texturized

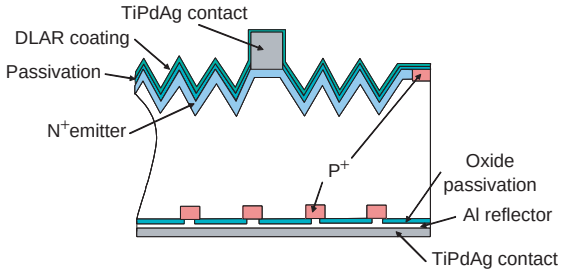


Figure 4.2.13: *Basic schematic of a Hi-Eta-Si cell.*

surface, an antireflective double layer (DLAR), passivation, and the back surface covered with a reflecting aluminum layer.

4.2.5.3 Multijunction Gallium Arsenide on Germanium Solar Cells (GaAs/Ge)

The various semiconductor layers of an MJ-GaAs/Ge cell are grown by epitaxial processes on a germanium wafer (Ge wafer) of 0.14 mm thickness .

Due to the sensitivity of reverse biasing, each single cell needs to be protected by an internal (integrated) or external **shunt diode.**

The preferred concept is an integral shunt diode which is realized by separating a small solar cell area (7 mm^2) by mesa-etching the rest of the cell. From this separated part, the top cell is etched away in order to reduce the forward voltage of the two remaining cells connected in series (Figure 4.2.15), namely the middle cell (MC) and the bottom cell (BC). The created so-called shunt diode is connected to the neighboring cell as a two-junction protection diode (Figure 4.2.16).

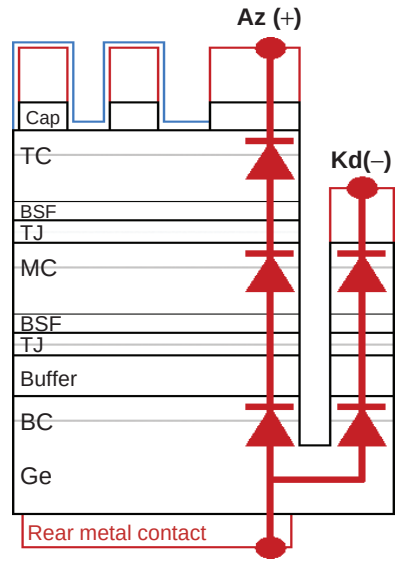


Figure 4.2.15: *Integral shunt diode.*

4.2.5.4 Electrical Characteristics of Solar Cells

Given the large variety of space-qualified solar cell types available from various manufacturers, only slight differences can be identified from the electrical characteristics of comparable cell technologies (e.g., SJ-Si, Hi-Eta or TJ-GaAs). For this reason the characteristics of only two examples, the Hi-Eta- and TJ-GaAs/Ge cells, are shown below in Tables 4.2.3 and 4.2.4.

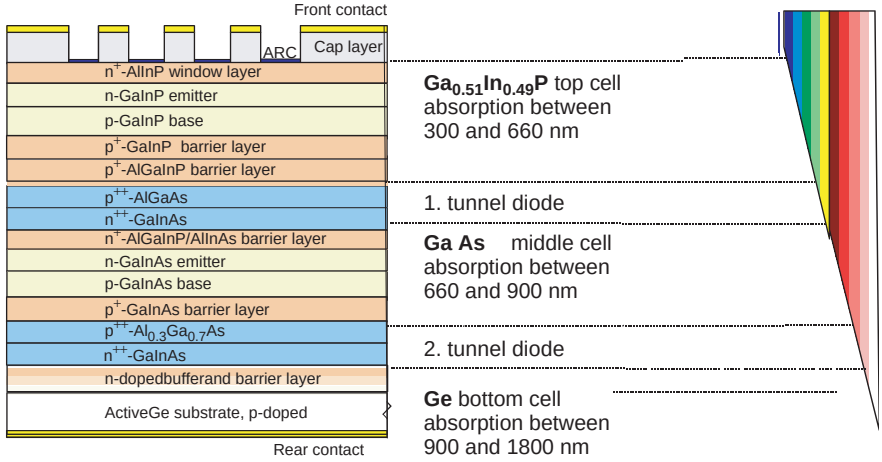


Figure 4.2.14: *Typical arrangement of the various layers of a TJ-GaAs/Ge solar cell consisting of three stacked single subcells with an indication of their light-absorbing spectral ranges.*

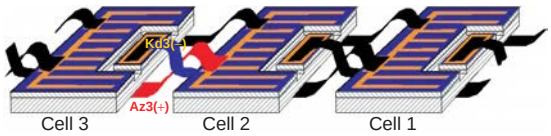


Figure 4.2.16: *Series connection of solar cells with integrated shunt diodes, e.g., anode of cell 3 [Az3(+)] (red conductor) is connected to cathode cell 2, and cathode of diode 3 [Kd3(-)] to anode cell 2 (blue conductor).*

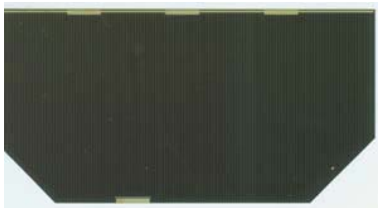


Figure 4.2.17: *Typical 8 × 4 cm² rectangular TJ-GaAs solar cell with cropped corners (shape optimized for least utilization of the Ge wafer area).*

Table 4.2.3: *Electrical characteristic of a Hi-Eta silicon cell.*

Electrical characteristics (ϑ = 28 °C)		Particle fluence		
	Nominal	3.0E+14	1.0E+15	3.0E+15
Parameter		***R***	***R***	***R***
V_{oc} [mV]	628	0.914	0.888	0.851
I_{sc} [mA/cm²]	45.8	0.882	0.846	0.758
V_{mp} [mV]	528	0.912	0.885	0.844
I_{mp} [mA/cm²]	43.4	0.876	0.837	0.757
Temperature coefficients				
$dV_{oc}/d\vartheta$ [mV/K]	−2.02	−2.14	−2.17	−2.20
$dI_{sc}/d\vartheta$ [mA/cm²/K]	0.030	0.045	0.055	0.059
$dV_{mp}/d\vartheta$ [mV/K]	−2.07	−2.22	−2.19	−2.25
$dI_{mp}/d\vartheta$ [mA/cm²/K]	0.022	0.023	0.027	0.035

4.2.5.5 Temperature and Radiation Effects

From known electrical characteristic such as those given in Tables 4.2.3 and 4.2.4, the temperature behavior and the radiation-dependent degradation effects on the *I*–*V* characteristic can be calculated for each cell type and, with that information, also the *I*–*V* curve for the entire SA. Including generally known loss factors, such as that for the cell cover glass ($K_{CG} \cong 0.99$),

Table 4.2.4: *Electrical characteristics of a TJ-GaAs cell.*

Electrical characteristics (ϑ = 28 °C)		Particle fluence		
	Nominal	3.0E+14	5.0E+14	1.0E+15
Parameter		***R***	***R***	***R***
V_{oc} [mV]	2660	0.957	0.946	0.934
I_{sc} [mA/cm²]	17	0.993	0.985	0.963
V_{mp} [mV]	2365	0.956	0.944	0.930
I_{mp} [mA/cm²]	16	0.990	0.979	0.955
Temperature coefficients				
$dV_{oc}/d\vartheta$ [mV/K]	−6.0	−6.4	−6.4	−6.5
$dI_{sc}/d\vartheta$ [mA/cm²/K]	0.01	0.011	0.012	0.014
$dV_{mp}/d\vartheta$ [mV/K]	−6.01	−6.7	−6.7	−6.9
$dI_{mp}/d\vartheta$ [mA/cm²/K]	0.009	0.01	0.011	0.012

Definition: The nominal values are referenced to the extraterrestrial AM0 solar spectrum (see Figure 4.2.11) at a cell temperature of 28 °C (301 K) and a solar constant value of 1371 W/m² valid for 1 AU (averaged Earth–Sun distance).

cell mismatch ($K_{CM} \cong 0.99$), parameter calibration ($K_{PC} \cong 0.97$), UV and micrometeorite effects (typically 0.25% per year), and taking seasonal solar intensity variations into account (winter solstice, WS = 1.034; summer solstice SS = 0.967), one can calculate the number of solar cells required to satisfy the BOL or EOL SA power demand. If a realistic cell coverage factor is applied, the required SA surface area can be calculated as well. Also the number of solar cells to be connected in series to form a string (voltage adjustment) and the number of strings to be connected in parallel to form an array can be calculated for all required current and voltage values or for a required operating point of the solar generator. Using a simplified model (following the equations of H.S. Rauschenbach) the array's operating point current (I_{op}) can be calculated as a function of the operating point voltage (V_{op}), and by varying V_{op} between 0 and V_{oc} one can obtain the *I*–*V* curve of the SA:

$$I_{op} = I_{sc} \cdot A_{cell} \cdot Np \cdot \left\{ 1 - Cb \cdot \left[e^{-\left(\frac{V_{op}}{Ca \cdot V_{oc} \cdot Ns} \right)} - 1 \right] \right\} \quad (4.2.6)$$

with

$$Ca = \frac{\frac{V_{mp}}{V_{oc}} - 1}{\ln\left(1 - \frac{I_{mp}}{I_{sc}}\right)} \tag{4.2.7}$$

and

$$Cb = \left(1 - \frac{I_{mp}}{I_{sc}}\right) \cdot e^{-\left(\frac{V_{mp}}{Ca \cdot V_{oc}}\right)} \tag{4.2.8}$$

Once the variables V_{mp}, V_{oc}, I_{mp} and I_{sc} have been supplemented with the corresponding temperature coefficients and the radiation dependent "remaining factors R" as shown in the following example for Ca, one obtains the temperature and radiation dependent I–V characteristic of the SA:

$$Ca(\vartheta, R) = \frac{\frac{V_{mp} \cdot R(V_{mp}) + dV_{pm}/d\vartheta \cdot (\vartheta - \vartheta_0)}{V_{oc} \cdot R(V_{oc}) + dV_{oc}/d\vartheta \cdot (\vartheta - \vartheta_0)} - 1}{\ln\left[1 - \frac{I_{mp} \cdot R(I_{mp}) + dI_{mp}/d\vartheta \cdot (\vartheta - \vartheta_0)}{I_{sc} \cdot R(I_{sc}) + dI_{sc}/d\vartheta \cdot (\vartheta - \vartheta_0)}\right]} \tag{4.2.9}$$

The equation for the I–V characteristic may then be written as

$$I_{op}(\vartheta, R) = \left[I_{sc} \cdot R(I_{sc}) + dI_{sc}/d\vartheta \cdot (\vartheta - \vartheta_0)\right] \times A_{cell} \cdot Np \times \left\langle 1 - Cb(\vartheta, R) \cdot \left\{ e^{\left[\frac{V_{op}}{Ca(\vartheta,R)\left[V_{op} \cdot R(V_{oc}) + dV_{oc}/d\vartheta \cdot (\vartheta - \vartheta_0)\right] \cdot Ns}\right]} - 1 \right\} \right\rangle \tag{4.2.10}$$

The generated electrical power of the array is $P_{op} = I_{op} \cdot V_{op}$.

The result can finally be multiplied by the previously mentioned degradation factors and the seasonal solar intensity factor, if these are known.

The generated electrical power of the array for summer solstice after five years of in-orbit operation and respecting the degradation factors is calculated as follows:

$$P_{op} = V_{op} \cdot I_{op} \cdot K_{CG} \cdot K_{CM} \cdot K_{PC} \cdot (1 - 0.0025)^5 \cdot 0.96 \tag{4.2.11}$$

Of course, there is also the possibility to calculate the reduction in SA power output caused by an angular deviation of the array from the ideal perpendicular orientation with respect to the Sun. The electrical power decreases with the cosine of the angular deviation:

$$P_{op}(\varphi) = I_{op} \cdot V_{op} \cdot \cos(\varphi) \tag{4.2.12}$$

At incident angles of $\varphi > 50°$ the refraction effect should be considered in the calculation.

The fact that less solar irradiation also reduces the operating temperature of the array is neglected in the formula used in Equation 4.2.9.

Conventions:

A_{cell} = Active area of the solar cell in cm^2

Np = Number of parallel solar cell strings

Ns = Number of series-connected solar cells per string

φ = Angle between the vector normal to the active PVA (Photovoltaic Solar Array) surface and the plane of the incident solar radiation.

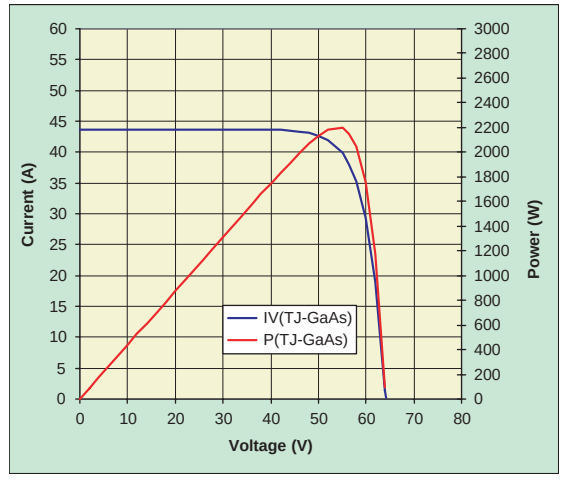


Figure 4.2.18: *Example of a calculated I–V characteristic for an array and the associated electrical power generated.*

Figure 4.2.18 depicts the calculated characteristic for the current, voltage and power of an SA with the following parameters:
$Ns = 25$
$Np = 100$
$A_{cell} = 26\,cm^2$
$V_{mp} = 2.277\,V$
$V_{oc} = 2.565\,V$
$I_{mp} = 0.016\,A/cm^2$
$I_{sc} = 0.0168\,A/cm^2$
AM0 (28 °C), BOL ($R = 1$).

In view of the future Solar Orbiter mission planned by ESA, the post-mission analysis of the long-term influence of temperature, radiation and micrometeorite effects on a solar generator in an extremely demanding environment (up to 165 °C solar cell surface temperature) for the evaluation of the Helios body mounted solar generator (German–US interplanetary probe, approaching the Sun as close as 0.3 AU) after 10 years in orbit (i.e., 20 aphelion and perihelion encounters) might be useful as a reference [4.2.23].

4.2.5.6 Solar Generator Technologies

Different spacecraft mission requirements may call for other SA configurations. State-of-the-art configurations are:

- Strings or arrays mounted directly onto the outer surface of the satellite body (**body-mounted arrays**).
- **Deployable flexible or rollout arrays**, for example motor-actuated deployment, like awnings.
- **Deployable rigid arrays** of rectangular panel shape.

Deployable flexible and rigid arrays are in most cases rotated 360° around their deployment axis. This allows the active PVA plane to be oriented perpendicular to the Sun at any time in order to acquire and convert the maximum possible solar radiation into maximum electrical energy.

For commercial high-performance SAs with an output power higher than 7 kW, a configuration with two deployable rigid arrays (wings) has so far been preferred, whereby each of the wings consists of four to six foldable panels. Satellites with an output power higher than 2 kW for EOL use exclusively rotating arrays. Earth observation and scientific (EOS) satellites often use a combination of body-mounted and deployable arrays.



Figure 4.2.19: *A 10 kW telecommunications satellite ARTEMIS (Source: ESA).*

The specific electrical power delivered by an SA as shown in Figure 4.2.19 is approx. 100 W/kg (BOL, AM0, 28 °C, 1 AU). The relation of electrical power to the required array panel area is approx. 280 W/m^2 (BOL, AM0, 28 °C, 1 AU).

4.2.5.7 Series Connection of Solar Cells

The number of solar cells to be connected in series (string length) is determined, as mentioned above, by the EPS architecture, or, to be precise, by the voltage operating point of the array as determined by the chosen topology of the primary power bus. The string length must be designed such that in the case of a DET-regulated array, the required **voltage operating point** (V_{op}) is slightly below (approximately 3–5%) the maximum power point voltage V_{mp} expected for EOL.

This measure safeguards a design uncertainty with respect to V_{op} (uncertainty in the analysis of the array operating temperature, drift of the electrical characteristics of the SA power regulation loop components, etc.), which could lead to a drift in the voltage operating point V_{op} beyond MPP in the direction of the open-circuit voltage, leading to an insufficient SA power provision.

Such a requirement must not be considered a design criterion for **MPPT regulation**. In the case of MPPT regulation the string length must be chosen

such that V_{mp} is, using step-down (buck) converters, always approx. 10% higher than the maximum bus voltage during the periods when power is delivered by the array, and always 10% lower than the minimum bus voltage in case of using step-up (boost) converters. The 10% voltage tolerance is needed for the linear behavior of the MPPT control loop.

4.2.5.8 Parasitic Capacitances and Inductances

The stability as well as the power loss of an SA power regulator can be strongly influenced by the electrodynamic output characteristic of the SA. The main role is played by the capacitive reactance of the (Coulomb) output capacitance at all operating points, while the dynamic resistance R_d becomes influential only at operating point voltages around V_{mp} and above. Knowing these parameters allows precise design of the power regulation as well as of the SA simulator to be used for satellite functional testing, since it is then well adapted to the real electrical output parameters of the SA.

4.2.5.9 Operating Temperatures

For the deployed SA, these are:

- LEO, Sun-synchronous dawn–dusk orbit (e.g., Sentinel-1): approx. 70 °C average for GaAs cells
- LEO, with significant Earth IR and albedo coupling (e.g., Aeolus): approx. 80 °C average for GaAS cells
- MEO (e.g., Galileo): 65 °C
- GEO (Telecom): approx. 60 °C.

For a body-mounted SA, these are:

- Roughly 20 °C above the values of deployed SAs, thus 90–100 °C for LEO missions
- In case of a large Earth view factor of the PVA (i.e., the Earth is completely in the field of view (FOV) of the PVA), also 115–120 °C may occur.

The values for silicon cells are about 10–15 °C lower.

4.2.5.10 Electrostatic Charging, Discharging and Discharge Effects

Charging of satellite structures during so-called "solar substorms" can lead to electrostatic potentials from several hundred to a thousand volts between adjacent surfaces, in particular during sunlit phases.

In the presence of a charged particle environment the satellite chassis is charged negatively at a rate of −5 V/s. The front surface of the SA faces the Sun, and charge from the usually non-conducting surfaces of the cell cover glasses is constantly bled off via photoemission, while other areas, such as the gaps between the solar cells, will remain negatively charged since they are shaded by the edges of the solar cells.

A potential difference between the cover glass and underlying cell of several hundred volts is gradually built up at a rate of approx. 3 V/s.

If the voltage difference reaches a sufficient level, **primary electric arc discharges** will occur. These discharges carry very little energy and are harmless as a single event. However, they can set free plasma from the molecular cell structure which settles in the gaps between the cells. Several hundred discharge events can lead to a plasma concentration establishing a low ohmic connection to the adjacent solar cell. If the difference between the nominal operating point voltages of the adjacent cells is high enough and if an appropriate photocurrent is generated within the cells, a **secondary sustained arc** could form carrying sufficient energy to produce a short circuit. The energy of a sustained secondary arc is usually high enough to cause permanent damage by evaporation of solar cell material and of the underlying insulation (string failures) [4.2.4], [4.2.5], [4.2.6].

Since the primary electric arc discharges can only be avoided by a conductive coating of the array surface, which entails other severe disadvantages, the conditions allowing secondary electric arcs must be eliminated.

This is achieved by appropriate measures in the solar cell string design. The goal is to make sure that the voltage difference between adjacent cells as a function of the gap size between the cell edges never reaches the discharge level and that the driving current remains low enough. The latter is achieved by adding a decoupling diode in series to each string and by parallel connection of the strings to an array behind the diode. Table 4.2.5 summarizes the analytically determined worst case boundary conditions for secondary sustained arcs.

Table 4.2.5: *Boundary conditions for secondary sustained arcs.*

String current [A]	Voltage difference Gap = 0.9 mm [V]	Voltage difference Gap = 0.7 mm [V]	Voltage difference Gap = 1.1 mm [V]
0.3	101.4	93.8	109.1
0.5	68.2	63.5	73.0
1.0	43.3	40.8	45.9
1.5	35.0	33.3	36.8
2.0	30.9	29.5	32.3
2.5	28.4	27.2	29.6
3.0	26.7	25.7	27.8
3.5	25.6	24.6	26.5
4.0	24.7	23.8	25.5
4.5	24.0	23.2	24.8
5.0	23.4	22.7	24.2
5.5	23.0	22.2	23.7
6.0	22.6	21.9	23.3
6.5	22.3	21.6	22.9
7.0	22.0	21.4	22.6

4.2.6 Energy Storage

Energy storage is an integral part of the EPS of a spacecraft. Every spacecraft relying for its primary energy only on solar energy needs some kind of energy storage to ensure the support of the electrical loads also during eclipse phases or other periods of low solar irradiation on the SA.

Energy storage is usually realized by a **battery**, although other energy storage systems like **flywheels** or **fuel cells** have also been sporadically used in some spacecraft.

A battery consists of individual electrochemical cells connected in series for multiplication of the cell voltage. Identical battery cells or cells with thoroughly harmonized parameters, such as lithium-ion cells, can be connected in parallel to enhance the amount of stored electric charge (capacity).

Batteries consist of primary or secondary cells.

In **primary cells** the chemical energy is converted to electrical energy. In contrast to secondary cells the conversion process in primary cells is irreversible. Since they cannot be recharged, their application is limited to short mission durations of several days to a few weeks. The most common primary cell types are silver–zinc, lithium thionylchloride, lithium–sulfur dioxide, lithium monofluoride and thermal cells.



Figure 4.2.20: *NiH_2 battery with 20 cells in series.*

A **secondary battery** allows some thousands of charge and discharge cycles depending on its depth of discharge (DOD), the charging method and the operating temperature. This makes it useful as an energy storage system for a spacecraft mission with an expected in-orbit lifetime of many years, such as LEO missions of seven years and more with 14 to 15 Sun and shadow phases daily. In such operations, the battery has to perform nominal for 15 (orbits) × 365 (days) × 7 (years), which equals 38 325 charge–discharge cycles.

4.2.6.1 Secondary Battery Technologies

Rechargeable batteries used for energy storage systems in spacecraft rely on three technologies: **nickel–cadmium** (NiCd), **nickel–hydrogen** (NiH_2) and **lithium-ion cells** (Li-ion). The two main technologies are the NiH_2 battery (Figure 4.2.20), which has been used for over two decades for space applications, and the new Li-ion battery (Figure 4.2.21) used increasingly since 2002.

For present spacecraft developments NiCd is usually not within the scope of application, and this is becoming more and more the issue also for NiH_2.

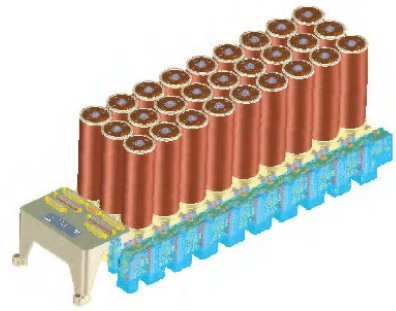


Figure 4.2.21: *Li-ion battery with nine cell packets in series and three parallel cells per packet.*

Performance and Life-Determining Parameters

Battery Life Expectancy

Secondary batteries have a limited life expectancy because of unwanted chemical and physical changes, like the loss of chemically active materials which irreversibly degrade the electrical performance of the battery [4.2.8].

The **calendar life** is the time in which the battery, or a certain part of its capacity, becomes useless [4.2.8], independent of whether the battery has been actively operated or not. This effect is significantly noticeable for Li-ion cells only.

The **cycle life** of a battery is defined as the number of charge–discharge cycles a battery can perform before parts of its nominal energy content fall below a usable battery voltage. This degradation of battery voltage is a gradual process caused mainly by the continuous increase of the cell's internal resistance [4.2.8].

The cycle life depends directly on the DOD that the battery is operated with during its operational life. For example, cycle lifetimes of some thousands of cycles for an 80% DOD and some hundreds of thousands of cycles for a 10% DOD are possible.

It is good practice in space applications to follow the rule of allowing a continuous DOD of 85% for NiCd and NiH_2 and 60–70% for Li-ion cells.

The cycle life for NiCd and NiH_2 cells (see Figure 4.2.22) can be considered to be equal. For Li-ion cells most manufacturers have developed an algorithm to calculate capacity losses (capacity fading) as a function of calendar and cycle life.

For space-qualified NiCd and NiH_2 batteries a good approximation of cycle life can be calculated with the formula

$$CL = 207\,800 \cdot e^{(-0.0272 \cdot DOD\,[\%])}. \tag{4.2.13}$$

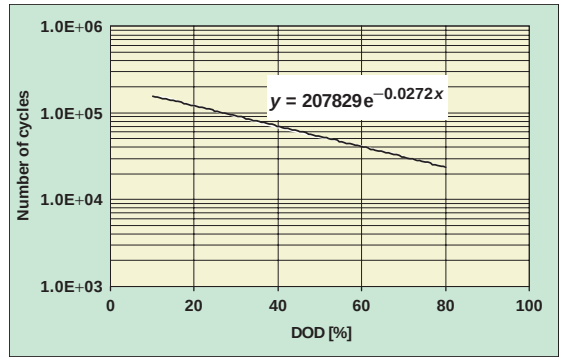


Figure 4.2.22: *Typical cycle life of NiH_2 batteries.*

The following mathematical relationship is in fact not applicable for all Li-ion cell types, but is an appropriate way to estimate the capacity fading (CF) of large cells with capacities of 20 ampere-hours and above:

$$CF\,[\%] = \text{Calender Loss}\,[\%] + \text{Cycling Loss}\,[\%]$$
$$CF = ka\sqrt{\text{time}} + kc\sqrt{\text{number of cycles}}, \tag{4.2.14}$$

with ka the calendar life in years:

$$ka\,(\text{temperature}/^{\circ}\text{C}) = 0.0009 \cdot \theta^2 - 0.0129 \cdot \theta + 0.1533 \tag{4.2.15}$$

$$kc\,(DOD) = 0.005 \cdot DOD\,[\%] + 0.021 \tag{4.2.16}$$

The **shelf life** is the time that an inactive battery can be stored before it becomes unusable, usually considered as being less than the nameplate capacity after a prescribed recharge cycling.

Temperature Effects

Increased operating and storage temperatures might reduce battery lifetime significantly.

The warmer the battery, the faster the chemical reactions. High temperatures can thus provide

increased performance, but at the same time the rate of the unwanted chemical reactions will increase, leading to a reduction in calendar life [4.2.8]. In addition, the negative temperature coefficient of the cell voltage at increased temperature causes a lower operating voltage. At constant power bus load and consequently at constant battery discharge power a reduction of the battery voltage increases the battery discharge current, and thereby also the DOD, since DOD is a function of the ampere-hours taken from the battery. As mentioned above, a higher DOD reduces the cycle life of the battery.

Depth of Discharge

The relation between cycle life and DOD is amply discussed above: the empirically determined mathematical algorithm is an **exponential function** with its exponent increasing with lower DODs.

This holds an important lesson for the battery designer: significant increases in cycle life can be achieved by a well considered thermal design.

Manufacturers of battery cells should either provide sufficient data for the analysis of the allowable number of cycles as a function of DOD, or, even better, provide the mathematical relationship itself.

Microcycles with a DOD below 0.1%, as usually occur with impulse loads, do not influence cycle life expectancy.

Cell Voltage Effects

All battery cell technologies have a characteristic **operating voltage range**. The definition of the usable voltage range is a consequence of the onset of undesirable chemical reactions which take place beyond the safe working range.

Once all active chemicals are converted to stored energy during the charging process, which means that the battery is completely charged, any further charge energy applied to a battery cell will heat up the cell. If overcharging happens over too long a time or too often, irreversible reactions will occur leading to permanent damage, particularly in Li-ion cells. The permanent increase of cell temperature by overcharging leads to increasing cell pressure, which means that uncontrolled overcharging can cause bursting or even explosion of a cell, releasing dangerous chemicals or even causing open fires [4.2.8]. Tests with Li-ion cells demonstrated that overcharging by increasing the cell voltage at the end of the charge process by 0.1 to 0.3 V causes a decrease in cycle life of up to 80%. Also, only a few deep discharges below the values specified by the manufacturer (typically 2.5 to 2.7 V) can cause permanent damage in Li-ion cells.

NiCd cells are somewhat less sensitive to overcharging. A permanent overcharging of 5–10% causes only a small temperature increase of less than 10 °C. The NiH_2 cell is even less sensitive to permanent overcharging of up to 20%. A charge factor (*k*-factor) of 1.2 is even used to balance the capacity of the individual battery cells (cell balancing).

Protective circuits and well adapted charge control methods must assure operation of the battery cells within their recommended operating ranges. A reasonable operational safety margin should be included in the design of a charge controller.

Memory Effect

The so-called "memory effect" is another manifestation of the changing morphology of the cell components with age. NiCd, and to a lesser extent NiH_2, cells "remember" how much constant energy was drained from the battery over many previous cycles, and only accept this particular amount of recharge energy. What happens in fact is that repeated shallow charges cause crystalline changes in the cell electrodes which increase the internal resistance of the cell. This is why, during charge as well as during discharge, the permitted voltage limits are reached much earlier, thus reducing battery capacity. Long, slow charging phases like "trickle charge" and increased operating temperatures support this undesirable behavior [4.2.8]. Periodic reconditioning of nickel-based battery cells allows recovery of the capacity lost by the memory effect. It would be even better, if possible, to cycle the battery, after having experienced a few hundred cycles at low DOD, for some 10 cycles with higher charge and discharge energy.

Reconditioning and Capacity Restoration

It is often possible to restore the full capacity by one or even several deep discharges with a low and controlled current down to a cell voltage of 1 V [4.2.8]. Reconditioning of Li-ion cells using deep discharge to a low cell voltage level is not allowed, as mentioned above.

It is often possible to restore a cell to, or near to, its full capacity essentially by repeating the formation process to break down the larger crystals into their previous smaller size.

Battery Charging

In most cases, a battery's total losses or early degradation are caused by improper charging methods. The use of a well-designed and adapted charge regulator is mandatory for secure und durable battery operations. Suitable charge regulator concepts are discussed in the following section.

Interactions between Battery Cells

Interactions between cells can occur in multicell batteries caused by unbalanced temperature conditions in individual cells or in conjunction with the spread in electrical characteristics caused by manufacturing tolerances. Nonuniform aging patterns can cause some cells to absorb less charge energy than others.

In a series connection of battery cells such effects would drive the "weak" cell with reduced capacity to reach the end-of-charge voltage prematurely, while the other cells with nominal characteristics would not yet have reached the completely charged state. If the charge status of the weak cell is ignored and more charge energy is applied, the weak cell will be overcharged and possibly damaged [4.2.8]. If the charge process is ended once the weak cell has reached its end-of-charge voltage, which by all means would be the correct charge process, then the battery's nominal capacitance would not be reached; however, individual cells would not be stressed excessively.

In analogy to the charge behavior also in the discharge mode, a weak cell reduces the available capacitance because it reaches its end-of-discharge voltage while the other cell could be discharged further. Continuation of the discharge will lead to deep discharge of the weak cell, and in the case of Li-ion cells very quickly damage the cell.

Premature Failure

The most likely cause of premature failure of a battery is abuse; that is, subjecting a battery to conditions for which it was never designed.

Besides physical abuse of a battery, the following operational conditions are considered as misuse [4.2.8]:

- **Increased discharge current:** The guideline for good cycle life is a continuous discharge current $\leq 1C$ and an impulse discharge current $\leq 2C$ (with C as a current value equivalent to the nominal (nameplate) capacity of the battery).
- **Increased charge current:** The guideline for good cycle life is a continuous full charge current $\leq 0.8C$ and a trickle charge current of $\leq C/100$ for NiCd and $\leq C/60$ for NiH_2.
- Too high or too low **operating temperatures**: The guidelines for operating temperatures are listed in Table 4.2.6.
- Use of a **charge regulator** designed for another cell technology.
- **Overcharge** or **deep discharge**.
- Excessive **vibrations** or **mechanical shock**.

Table 4.2.6: *Temperature ranges for optimal battery operation.*

Temperature regions	NiCd and NiH_2	Li ion
Operating	–10 to +25 °C	–10 to +35 °C
Non-operating	–15 to +35 °C	–20 to +40 °C
Charge optimum	–10 to +10 °C	+15 to +25 °C
Discharge optimum	+10 to +25 °C	0 to +30 °C

Cell Balancing

As already discussed in the section describing cell voltage effects, Li-ion cells do not tolerate overcharge and deep discharge outside the voltage limits specified by the manufacturer. As described above (see interactions between battery cells) unequal properties of individual cells might lead to those operational conditions causing possible interactions of individual cells among themselves. To avoid this effect for Li-ion cells, the so-called method of **cell balancing (charge balancing)** is used. The principle is described in Section 4.2.6.4.

For batteries connected in parallel the requirement for cell balancing is not that important since batteries connected in parallel have the tendency to balance themselves (self-balancing).

Selection of Cells

In order to avoid problems caused by manufacturing tolerances, cells must be combined with a battery following specific selection criteria, such as only using cells from the same manufacturing batch, or assuring minimal electrical characteristic deviations by classification using tests [4.2.8].

4.2.6.2 Comparison of Battery Technologies

A comparison of battery technologies is based on physical proportions and electrical properties. Besides the differing performance characteristics, the different chemical cell compositions also require different operating conditions. The most important parameters were discussed above.

The NiH_2 cell, characterized by robust and stable chemical composition, not only allows deep discharge and overcharge (as long as the heat is dissipated sufficiently and the internal cell pressure of up to $70 \cdot 10^3$ hPa is continuously monitored), but also tolerates reverse current associated with polarity reversal of the cell voltage. For the time being, the NiH_2 system is the one with the most in-orbit operational experience and has demonstrated excellent cycle lifetimes. The NiCd cell can be judged similarly, but with lower overcharge and less tolerable reverse current.

The Li-ion cell must be protected from all these operational modes and therefore requires a sophisticated **battery management system**. Also, the relatively small amount of in-orbit operational experience must be compensated by higher protection efforts. Despite the cost impact of the additional effort to guarantee safe in-orbit operations, the Li-ion cell offers, with its:

- more than double energy density properties (watt hours per weight, Wh/kg)
- relatively low voltage variation between full charge and discharge states
- high charge efficiency, low power losses (little heat generation) and
- small volume,

many good arguments for its present very dominant use in spacecraft.

Guidelines for Li-ion Cells: Since there is not yet enough in-orbit operating data available, the Li-ion cell should not be operated with more than 30% DOD for missions requiring a high number of cycles (LEO missions might last several years) and currently with not more than 70% DOD for all GEO missions.

Energy storage systems for small satellites in particular require small volumes with low weight. NiH_2 technology has experienced many new developments and improvements in energy density during its use in space applications. The development steps led from single cells in separate pressure containers (individual pressure vessel, IPV) via double-cell pressure containers (common pressure vessel, CPV) to pressure containers comprising 22 to 28 integrated cells (single pressure vessel, SPV). All those improvements could not compensate the decisive advantage of the Li-ion cell technology of providing a nominal (average) cell voltage of 3.6 V (in comparison to 1.25 V for NiH_2 and 1.2 V for NiCd). Tables 4.2.7 and 4.2.8 compare the properties on the cell and battery level.

In contradiction to NiH_2, space-qualified Li-ion battery cells of individual manufacturers develop different charge and discharge characteristics. The most important parameter in this regime, namely the cell voltage as a function of the battery state of charge (SOC), and also versus DOD, is quite different, as shown in Figures 4.2.23 and 4.2.24.

4.2.6.3 Principles for Battery Capacity Sizing

Before the required battery capacity can be calculated for a given spacecraft mission the required stored energy must be known for all mission dependent operational modes, taking anomalies and battery degradation due to mission lifetime into account.

It is assumed that the total primary power bus power (P_{bus}) for the longest possible battery discharge duration (t_{dch}), usually the eclipse phase during one orbit, is known either by summing all single constant loads or by averaging load power profiles, say, by using a spreadsheet program. The power losses in power distribution due to voltage drops along the cables (P_d) as well as the energy conditioning losses within the EPS (P_C) must be determined as well.

Further, the mission duration and the orbit parameters must be known, allowing calculation of the average eclipse duration (t_{ae}) and the number of battery discharges due to the number of eclipses (N_E) during the time of the mission.

Table 4.2.7: *Cell technology and energy density comparison.*

Parameter	NiCd	NiH_2 IPV	NiH_2 CPV	NiH_2 SPV (#22)	Li ion
Specific energy density [Wh/kg]	37	44	51	60	135
Charge efficiency η_{ch} [%]	0.83	0.83	0.835	0.85	0.95
Nominal voltage $V_{c\text{-nom}}$ [V]	1.2	1.25	2.5	28	3.6
End-of-charge voltage Vc_{EOC} [V]	1.5 at 10 °C	1.55 at 10 °C	3.1 at 10 °C	34.0 at 10 °C	4.1 at 20 °C

Table 4.2.8: *Comparison of battery technologies.*

Parameter/criteria	NiCd	NiH_2	Li ion
Specific energy density	28–32 Wh/kg	35–50 Wh/kg	70–110 Wh/kg
Specific volume	14–18 l/kWh	22–30 l/kWh	8–9 l/kWh
Cycle life vs. DOD	Good	Excellent	Satisfactory (degradation is dependent on the number of cycles and the lifetime)
Discharge voltage behavior	Satisfactory	Good	Excellent
Battery management effort	Nominal (=100%)	Reduced (=60%)	High (=150%)
Recurring cost per Wh	170	€300 (IPV) €220 (CPV) €200 (SPV)	€60–100

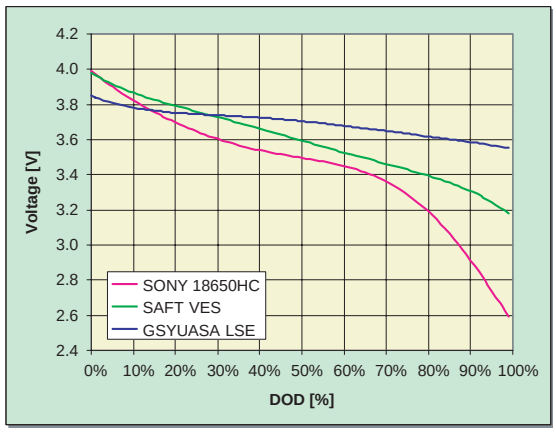


Figure 4.2.23: *Discharge voltages as a function of DOD at C/2 charge*

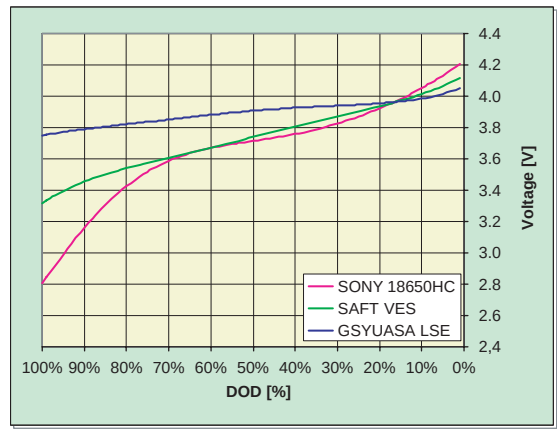


Figure 4.2.24: *Charge voltages as a function of DOD at C/2 charge*

Before the allowed DOD can be established (for NiCd or NiH_2 with Equation 4.2.10 and for Li ion according to the guideline of 30% in LEO and 70% in GEO), an analysis of whether, and if so how many, additional battery discharges will occur due to load peaks on the bus during the sunlit phase must be undertaken.

In addition, a cycle life margin of at least 30% has to be planned for. This means that for NiCd and NiH_2 batteries the allowable DOD is calculated as $N_E/0.7$ as long as no other discharge cycles occur during the sunlit phase, while in the case of Li-ion batteries the term DOD has to be divided by 0.7.

Calculation Process

The total power demand of the spacecraft is

$$P_R\ [W] = P_{bus} + P_d + P_C \qquad (4.2.17)$$

The battery energy required is

$$E_{\text{dis}}\,[\text{Wh}] = P_{\text{R}} \cdot t_{\text{dch}}\,[\text{s}]/(3600\,[\text{s}] \cdot \eta_{\text{ch}}) \qquad (4.2.18)$$

with η_{ch} as the discharge efficiency, such as for a BDR within an EPS providing a regulated bus (BR). Of course in Equation 4.2.18 time intervals other than seconds can also be used.

The required nominal battery energy is

$$E_{\text{bat}}\,[\text{Wh}] = E_{\text{dis}}/(DOD\,[\%]/100) \qquad (4.2.19)$$

With E_{bat} the necessary nominal capacity ($C_{\text{b-nom}}$) of the battery must be calculated. For that, the average battery discharge voltage during discharge periods must be known. To a good approximation it can be calculated by multiplying the number of battery cells connected in series (Ns) with the nominal cell voltage (as given in Table 4.2.7). Ns is the rounded integer of the quotient of the required maximum battery voltage ($V_{\text{b-max}}$) with the maximum cell voltage at the end of the discharge cycle (Vc_{EOC}). $V_{\text{b-max}}$ in turn directly depends on the chosen EPS architecture and the defined bus voltage.

For the number of battery cells connected in series:

$$Ns = \text{integer}\,(V_{\text{b-max}}/Vc_{\text{EOC}}) \qquad (4.2.20)$$

The nominal capacity of the battery $V_{\text{c-nom}}$ from Table 4.2.7 is

$$C_{\text{b-nom}} = E_{\text{bat}}/(Ns \cdot V_{\text{c-nom}}) \qquad (4.2.21)$$

If an Li-ion battery is to be used, the calculated capacity fading (CF) using Equations 4.2.14–4.2.16 must be respected to determine the real BOL battery capacity $C_{\text{b-BOL}} = C_{\text{b-nom}}/(1 - \text{CF}\,[\%]/100)$ to be installed.

4.2.6.4 Battery Charge Control

Nickel–Cadmium and Nickel–Hydrogen Batteries

Voltage–Temperature Method (V–T Method)
This method applies stepwise selectable, battery temperature and charge-current-compensated end-of-charge voltage (EOCV) limits for charging the battery. Figure 4.2.25 shows a design example for an NiH_2 battery with 16 cells in series. The compensation of

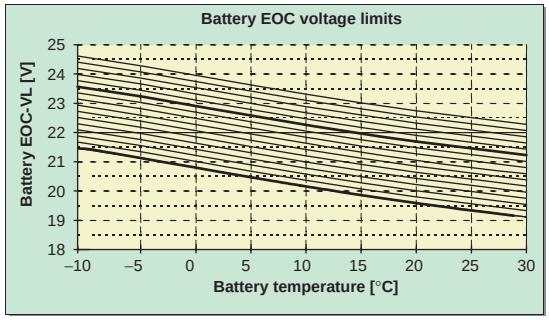


Figure 4.2.25: *Adjustable V–T end-of-charge voltage levels.*

the cell voltage temperature coefficient (TC) is absolutely necessary for NiCd and NiH_2 batteries, since negative TC cannot be neglected in order to achieve complete charging of the battery without permanent control of the charge voltage and to avoid permanent overcharge. Charge current compensation ensures compensation for the voltage drop caused by the charge current through the battery, whose internal resistance increases with the number of cycles over the battery life. If not compensated, the increasing charge-current-dependent voltage drop on the battery would lead to a premature EOCV at a consistent (uncompensated) charge voltage.

Typical values for the increase of internal resistance for long mission durations are between 15 and 30%. Considering that batteries with high capacities or with a low number of cells connected in series have an internal resistance of approximated 20–40 mΩ, the charge current compensation can be abandoned for charge currents up to approximately 30 A.

Ampere-hours Integration
The product of discharge current (I_{dch}) and discharge time (t_{dch}) is compensated by the product of charge current (I_{ch}), charge time (t_{ch}) and charge factor. The charge factor (k) can be selected as a step function:

$$I_{\text{dch}} \cdot t_{\text{dch}} = k \cdot I_{\text{ch}} \cdot t_{\text{ch}} \qquad (4.2.22)$$

Cell-Pressure-Dependent Charge Control (for NiH_2 only)
The linear proportional dependency of the internal cell pressure on the SOC is measured either by a pressure indicator or by a resistive strain gauge and transformed into a temperature-compensated

analog signal indicating the actual SOC of the cell or the whole battery. This method is well suited for the charge control of batteries with several hundred up to several thousand charge–discharge cycles without requiring in-orbit recalibration.

Lithium-Ion Battery

For charge control of Li-ion batteries with cells not requiring periodic cell balancing, a charge voltage and charge current controlled concept similar to that for NiCd and NiH_2 batteries can be used. In this case a highly precise voltage measurement for each single cell or for each parallel cell connection is required to avoid any overcharge (see also the discussion on cell voltage effects in the previous section). As soon as the first cell reaches the allowable EOCV, the charge current must be reduced so that the allowed charge voltage is constant for each cell and is not exceeded at any time (taper charge principle). Temperature as well as charge current compensation should rather not be used since very small adaptation errors could lead to stressing overloads.

If cell balancing is required it should be performed just before every eclipse season (GEO) or every four to six months during battery recharge after an eclipse. This requirement leads to a more complex charge control approach for Li-ion batteries.

4.2.7 Design Fundamentals of EPS Systems

4.2.7.1 Bus Voltage

If no specific requirement for a defined bus voltage exists, the state-of-the-art main bus voltages described in Section 4.2.4 should be used as a guideline because for these voltage ranges much OTS equipment is available. In any case, certain under- and overvoltage criteria must be met for the main bus.

For the regulated bus all non-essential loads should be disconnected if an undervoltage of more than 10% of the allowed value is reached. If fuses are used in the power distribution interfaces to protect the main bus, delay times of at least 50 ms before initiation of load shedding at undervoltage conditions must be considered in the design, in order to grant enough time for fuse blowing in case of a short circuit.

In the case of the unregulated bus all non-essential loads should be shed at a battery voltage corresponding to a charge state of the battery sufficient to supply all essential loads until a safe spacecraft operational mode (safe mode) is reached.

Another often specified requirement is that a main bus should have the capability to recover automatically after a temporary or long-duration voltage **shutdown**, as soon as the cause of the shutdown has been removed and enough SA power is available.

4.2.7.2 Solar Array Power Conditioning

The basic functions of DET and MPPT as the leading methods for power transfer from the SA to the main bus were described in Section 4.2.4, as well as their suitability for the various bus alternatives. Therefore, only the advantages and disadvantages are described here.

Direct Energy Transfer

Advantages

DET electronics are simple and save weight and cost.

Disadvantages

The installed SA power must provide enough margin, because:

- The bus voltage and the actual battery voltage determine the operating point on the *I*–*V* curve of the SA sections which are connected directly (without BCR) to the battery for the purpose of energy storage. For the unregulated bus this is the entire SA. In particular at low battery voltage conditions (e.g., after battery discharge), not all available SA power can be transferred since a low battery voltage clamps the SA voltage operating point in the direction of low operating voltages. As the SA operating point in the DET operational mode is adjusted in the region of the SA short-circuit current I_{sc}, the decrease of the operating point voltage will reduce the useful SA power proportionally. Only with increasing battery charge voltage will the useful SA power increase until the EOCV is reached.

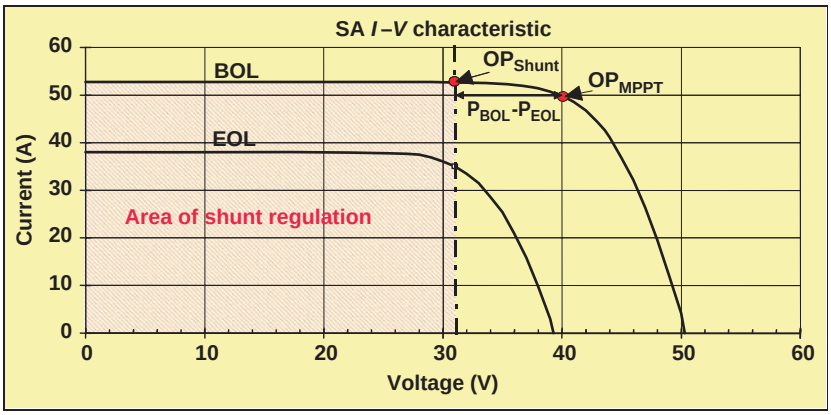


Figure 4.2.26: *Power adjustment and accommodation at a DET-controlled SA.*

- The SA voltage and the power operating point must be adjusted and optimized to EOL operational conditions, thus wasting part of the higher BOL power performance, as illustrated in Figure 4.2.26 by $P_{BOL} - P_{EOL}$.
- Because of the optimization of the operating point, no flexibility is given for the SA's cell string length (the reasons are described in detail in the section entitle "Series Connection of Solar Cells" in Section 4.2.5.7).
- Temperature changes as well as incorrectly analyzed operating temperatures will also lead to deviations from the optimum SA operating point.
- The function and control loop stability of the regulation system can be strongly influenced by the electrodynamic output characteristic of the SA, whereby the parasitic output capacitance of the SA plays a dominant role. Charge commutation across the parasitic SA capacitances, induced from control of SA power by shunt switching, may produce a response time delay in the control loop, and may effect increased switching transients and losses.

Maximum Power Point Tracking (MPPT)

Advantages

The advantages are as follows:

- Since with this method the MPP of the SA is automatically extracted, the maximum available SA power can be utilized whenever needed, independent of the SA voltage, current and temperature. This makes it possible to save between 8 and 25% of the installed SA power.
- MPPT introduces excellent flexibility with respect to cell string layout and string length, cell size and cell type.
- MPPT might be the solution for missions with strong variations in solar intensity (interplanetary missions).
- **Pulse width modulated** (PWM) DC/DC converters used for MPPT regulation are largely insensitive to the electrodynamic behavior of the SA. This aspect becomes more and more a focal point because the SA output capacitances are becoming increasingly higher through the use of megajoule solar cells (significantly fewer cells in series form one string), the development of ever thinner semiconductor layers for solar cells, and the permanently growing demand on higher SA power for high-power spacecraft.

Disadvantages

MPPT regulation requires more **circuit design effort** than DET, causing higher weight and cost.

For a rough comparison of the weight increase between MPPT and DET the following rules of thumb can be used:

- MPPT-regulated power transfer: 2.5 g/W (at 28 V output voltage), 2.4 g/W (at 50 V), 2.2 g/W (at 100 V) but offering 20% reduction by new developments.
- DET-regulated power transfer: $42\,g \cdot P_{SG}(BOL)/V_{bus}$ with P_{SG} as the maximum SA power to be transferred.

Of course, the additional effort for the MPPT electronics must be compared to the weight and cost savings for the SA associated with MPPT regulation:

- Body-mounted SA, electronic components only (without structure and mechanical systems): 5.5–6 g/W.
- Rotating SA wings, including all mechanical components: 25 g/W ($P_{SG} < 1200$ W), 17 g/W ($P_{SG} > 5000$ W).

The costs for the electronics per weight unit (approx. €70 000–90 000 per kg) and for the SA power per watt (approx. €800 per W) are already known.

4.2.7.3 Voltage Converters and Power Regulation

Every voltage converter and power regulator, such as those applied for DET and MPPT (solar array regulator, SAR), battery charge regulators (BCRs) and battery discharge regulators (BDRs), produces power conversion losses as a consequence of its **efficiency** (η), which must be taken into account when sizing SA power and battery energy. If for example a battery provides the required energy to the bus via a BDR, the battery has to provide the conversion losses within the BDR in addition to the bus power P_{BUS}. Therefore, $P_{BAT} = P_{BUS}/\eta$ is the total battery power to be provided. The block diagram in Figure 4.2.27 depicts the various **regulation options** within an EPS architecture and Table 4.2.9 lists the associated typical efficiencies.

For **buck and boost regulators** in secondary voltage converters – these are converters with galvanic separation (transformer coupling) between the power input and power output circuits – the values in the table can be used as well, by subtracting 2–3%.

Additionally, the regulator design has to take regulation loop stability and minimization of electromagnetic disturbances into account.

4.2.7.4 Grounding Concept

Generally the so-called **distributed single-point grounding** (DSPG) system is used in spacecraft. In this system all bus return cables from the users are connected to a common grounding point (CGP) at the negative potential of the main bus. The CGP is connected to the closest possible point on the electrical conductive structure of the spacecraft using a very short conductive connection. Deviations from this concept are only allowed for high-frequency equipment, since in this case the return line for the power supply has to be connected directly to the electrical conductive housing of the equipment for screening purposes.

Table 4.2.9: *Efficiencies for various power regulators.*

Regulation type	η at 20 V	η at 30 V	η at 40 V	η at 50 V	η at 70 V	η at 100 V
SAR						
Series regulation	0.940	0.955	0.965	0.970	0.975	0.980
Shunt regulation	0.940	0.960	0.970	0.977	0.982	0.988
MPPT control	0.930	0.941	0.952	0.960	0.966	0.970
BCR/BDR						
Buck regulator	0.900	0.920	0.935	0.950	0.960	0.970
Boost regulator	0.870	0.910	0.930	0.945	0.960	0.970
BCR						
Shunt regulation	0.920	0.940	0.955	0.965	0.975	0.980

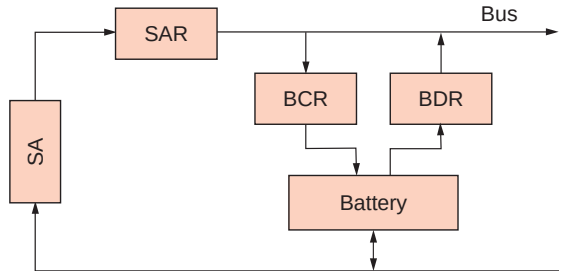


Figure 4.2.27: *EPS power regulator options.*

4.2.7.5 Measures for Power Bus Protection

Just as the power users have to be protected against under- or overvoltages on the primary power bus with appropriate measures, the bus must likewise be protected against overloads or short circuits caused by a user. To achieve this each single power line is protected either by a **passive protection element** like a fuse, or by an **active electronic current limiter**. Such

active devices are known as LCLs (Latching Current Limiters), FCLs (Fold-back Current Limiters) and SSPCs (Solid-State Power Controllers).

Application of Fuses

Generally the application of fuses for protecting power lines is a delicate matter requiring careful selection of fuse types and a thoroughly accomplished analysis of fuse current rating. For example, in ESA programs fuses are to be avoided whenever possible.

Use of Active Current Limiters

The **LCL** is a semiconductor switch which can be commanded in the on- and off-state. Its output overcurrent protection is set for a defined time span to a defined value (typically approx. 160% of the nominal current for 5–10 ms).

If the current increases above these limits the load will be disconnected. The LCL will stay in the off-condition until it is reactivated by an on-command.

The **FCL** is a semiconductor switch which cannot be commanded. The switch is able to tolerate an overcurrent of 10–25% of the rated value for an unlimited time. If this current limit is exceeded, the FCL output current will be reduced to a lower value at a lower output voltage, but does not disconnect the user. As soon as the overload condition is removed, the FCL automatically reverts to its nominal operating conditions.

The **SSPC** is a commercial circuit similar to the LCL function. Its use in spacecraft is usually prohibited by the lack of space flight qualification.

4.2.7.6 Power Distribution

The following rules for power distribution within spacecraft are generally followed:

- Users supplied by the main bus are connected individually. This means that also redundant equipment or functions with their own auxiliary power supplies have to be connected via separate power lines.
- Each individual power line is distributed and twisted together with a dedicated return line of the same size.
- The power harness is designed to distribute all primary and secondary power with minimum degradation and permissible losses; this is also true under peak load conditions.
- Power return lines of users working with signal frequencies of less than 10 MHz must be returned directly to the CGP.
- The primary power distribution design within the spacecraft should include not less than two single commandable on/off switches for a single supply interface. The on/off switches are preferably semiconductor switches, or as an alternative current impulse relays (latch relays).
- If a relay is used, the overcurrent protection device must be connected between the relay contact and its connection to the main bus.
- The operational current density through a relay contact is only 50% of the rated current for the relay contact.

Bibliography

[4.2.1] ECSS-E-20A. *Space Engineering, Electrical and Electronic.*

[4.2.2] ESA PSS-02-10, Vol. 1. *Power Standard.*

[4.2.3] ESA PSS-02-10, Vol. 2. *Rationale for the Power Standard.*

[4.2.4] Mank, J. H. Spacebus 3000B Solar Array ESD Test & Evaluation Report. *SB3R-DSS-TR-1430-0001*, Friedrichshafen: EADS Astrium.

[4.2.5] Mank, J.H. Possible Anomalies on Solar Arrays due to Electrostatic Charging in GEO. Presentation to ESTEC, Noordwijk, February 2, 1999.

[4.2.6] Katz, I., Davis, V.A., Snyder, D.B. Robertson, E.A. ESD Initiated Failures on High Voltage Satellites. 6th SCTC (Spacecraft Charging and Technology Conference), 1998, pp. 29–42.

[4.2.7] ASTM G173-03. *Table: Extraterrestrial Spectrum.*

[4.2.8] M Power Solutions Ltd, www.mpoweruk.com/life.htm, 2005.

[4.2.9] Cosculluela, V., King, S., Roeder, R. Engineering Handbook for Electrical Power Systems, *ENS.05.00033.ASTR.* Astrium, Earth Observation, Navigation & Science.

[4.2.10] La Roche, G. *Solargeneratoren für die Raumfahrt: Grundlagen der photovoltaischen Solarkollektoren-Technik für Raumfahrtanwendungen.* Braunschweig: Vieweg Verlag, 1997.

[4.2.11] Patel, M.R. *Spacecraft Power Systems.* Boca Raton, FL: CRC Press, 2004.

[4.2.12] Thekaekara, M.P. *Solar Energy Outside the Earth's Atmosphere.* Oxford: Pergamon Press, and *Solar Energy,* **14**, 109–127, 1973.

[4.2.13] Bogus, K. Solar Constant, AMO Spectral Irradiance and Solar Cell Calibration. ESA Technical Memorandum *TM-160*, 1975.

[4.2.14] Fröhlich, C., Wehrli, C. *Reference Extraterrestrial Spectral Irradiance Distribution*. Davos: World Radiation Center.
[4.2.15] Fröhlich, C. *Contemporary Measures of Solar Constant: The Solar Output and its Variations*. Boulder, CO: Associated University Press, pp. 93–109, 1977.
[4.2.16] NASA. Space Vehicle Design Criteria (Environment). *Technical Report No. SP-8005*, 1980.
[4.2.17] Hyder, A.K. *et al. Spacecraft Power Technologies*. London: Imperial College Press and Sigapore: World Scientific, 2003.
[4.2.18] Larson, W.J., Wertz, J.R. *Space Mission Analysis and Design, Space Technology Library: Spacecraft Subsystem Power*, Third Edition, 1998.
[4.2.19] Berlin, P. *Satellite Platform Design*, Fourth Edition. Lulea and Umea: Department of Space Science of the University of Lulea and Umea, 2005.
[4.2.20] Griffin, M.D., French, J.R. *Space Vehicle Design*, AIAA Education Series. Reston, VA: AIAA, 1991.
[4.2.21] Messerschmid, E., Fasoulas, S. *Raumfahrtsysteme*, 2. Auflage. Heidelberg: Springer Verlag, 2004.
[4.2.22] Hallmann, W., Ley, W. *Handbuch der Raumfahrttechnik*, 2 Auflage. Munich: Carl Hanser Verlag, 1999.
[4.2.23] Kehr, J. Das Langzeitverhalten von Energieversorgungssystemen am Beispiel Helios und Symphonie,.Dissertation, Lehrstuhl für Raumfahrttechnik, Technische Universität München, 1987.

4.3 Thermal Control

Andreas Kohlhase and Reinhard Schlitt

4.3.1 Introduction

The task of a thermal control system is essentially to ensure that the temperatures of all mechanical, electrical and electronic units in a spacecraft are within **specified operating temperature ranges** during all mission phases. The technologies and techniques applied to reach this aim are selected according to the applicable temperature ranges, which according to [4.3.1] can be defined as follows:

- < 200 K cryogenic range
- 200 to 470 K conventional range
- > 470 K high-temperature range.

Known thermal processes have varying importance in these temperature ranges. Heat exchange by conduction and radiation should be minimized in the **cryogenic range** in order to reduce heat transfer into these low-temperature systems. Typical applications are optical systems in the infrared region or rocket stages with liquid oxygen and hydrogen as the propellant. Low heat conduction is also important in **high-temperature systems** (e.g., reentry bodies). In addition, heat protection systems based on ablation are applied: a protection shield consumes itself through evaporation at the hot surface with the effect that the released heat of evaporation is not introduced into the protection system.

Most of today's space systems operate in the **conventional temperature range**, since the employed electronics are developed and qualified for temperatures between about −40 and +50 °C. Science satellites, Earth observation systems and communication satellites as well as planetary and interplanetary probes and landing systems fall in this category. The mentioned operating temperature range applies to units which are located inside the spacecraft. For external units quite large ranges are defined (e.g., −170 to +90 °C for antennas and −160 to +80 °C for solar arrays).

The next sections address in detail the basic thermal principles, the development steps of analysis and testing, as well as examples of thermal technologies. The section closes with a description of an implemented design example including a typical operating sequence for a thermal control system.

4.3.2 Basic Thermal Principles

4.3.2.1 Environmental Conditions

The following explanations concern satellites in an Earth orbit but are analogously true also for orbits around other planets.

Only conditions outside the atmosphere are of relevance for satellites, since bodies also in low orbits interact with only a reduced number of atoms and molecules in the remaining atmosphere. For a thermal design the influence of **convective heat transfer** can therefore be neglected.

A satellite therefore interacts with its environment only by **radiation**, and the desired temperatures can be achieved by skillful control of heat fluxes within

the satellite and from the satellite to the environment. Components are able to absorb or give off heat by radiation and conduction. Heat exchange by radiation can be influenced by selection of suitable **thermo-optical surfaces** and heat transfer by conduction through materials with tailored heat conductivity. Moreover, **thermal contact resistance** plays an important role for heat conduction in multiple component systems.

The (waste) heat released in a spacecraft must be radiated to the space environment, which exhibits a background temperature of about $T_H = 3\,K$. For this purpose **radiators** are attached to the outside surfaces of a satellite; they have thermo-optical properties with low emissivity in the infrared range and low absorptance over the entire solar spectrum. In addition to the direct solar radiation on satellites in Earth orbits, the **absorbed environmental energy** due to solar planetary reflection and the infrared radiation of Earth as a warm body must be taken into account (Figure 4.3.1).

The **energy balance** of a satellite can therefore be illustrated in a simplified way as in Figure 4.3.2.

Depending on whether the external temperature is higher or lower than the satellite temperature, the insulation heat leak is either a heat gain or heat loss for the satellite. A physical equilibrium temperature is established when the sum of the received energy, including the thermal dissipation of satellite components, equals the energy which is radiated into the environment.

Thermal control guarantees specified working temperatures for all components of a satellite and is therefore a service task for the remaining subsystems. The relevant interfaces to components, units and to the satellite system are exemplarily listed in Figure 4.3.3.

The solar energy received near Earth is expressed by the **solar constant** S and is according to [4.3.2] defined as the radiation which impinges on a surface per unit of time with the surface element arranged perpendicular to the radiation and located outside the atmosphere at a distance of 1 AU to the Sun. Since the Earth revolves around the Sun in an elliptical orbit, the value of the solar constant is not constant throughout one Earth orbit [4.3.3]. Depending on the season, the global solar energy received is calculated for the Earth as follows:

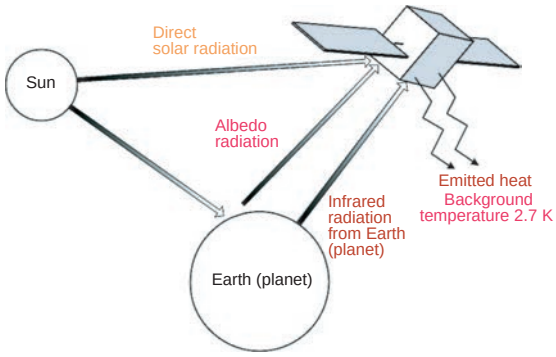


Figure 4.3.1: *Thermal radiation environment for a satellite in a near-Earth orbit.*

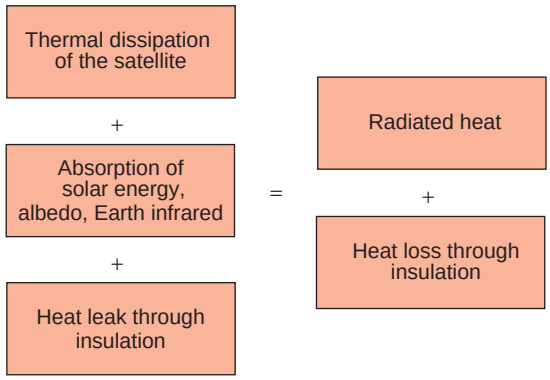


Figure 4.3.2: *Energy balance of a satellite in a low Earth orbit.*

$$S = S_0 \left[1 + 0.033 \cos\left(360° \frac{n}{365} \right) \right] \qquad (4.3.1)$$

where:
n = day of the year (n = January 1 → 3),
S_0 = 1371 W/m^2, the average present solar constant based on 1 AU (149 597 870 691 m ± 30 m).
The maximum radiation at perihelion (January 3) is therefore 1428 W/m^2 and the minimum radiation at aphelion (July 4) is 1316 W/m^2.

The solar radiation reflected from the Earth and impinging on the satellite is called **albedo** and amounts to 30% of the solar radiation, based on the total spectral distribution of the solar radiation. Albedo radiation should be considered for cases where part of the Earth's surface is illuminated. The radiation can temporary vary between 5 and 60%, depending

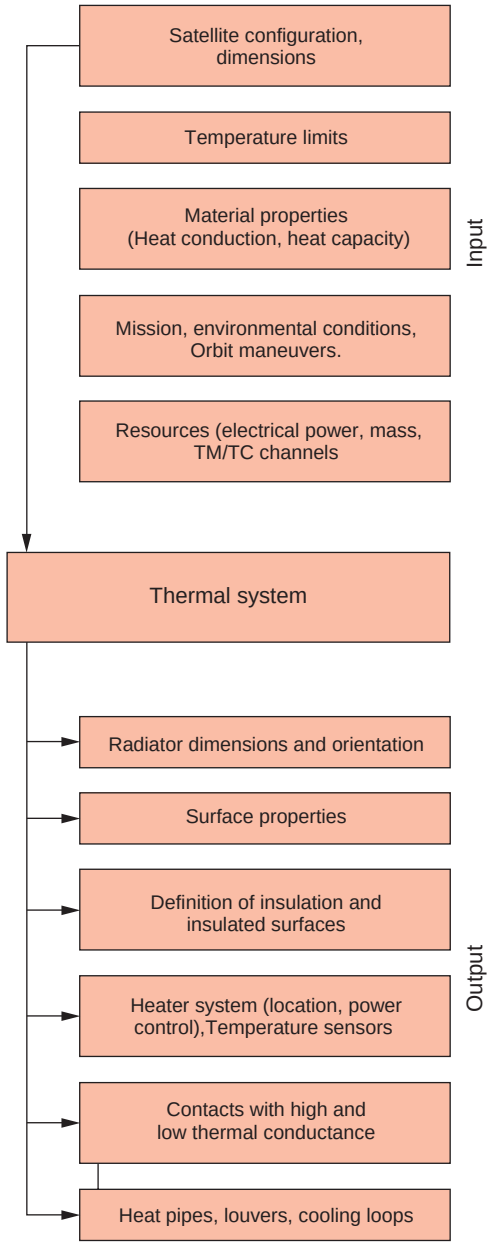


Figure 4.3.3: *Interfaces of thermal control.*

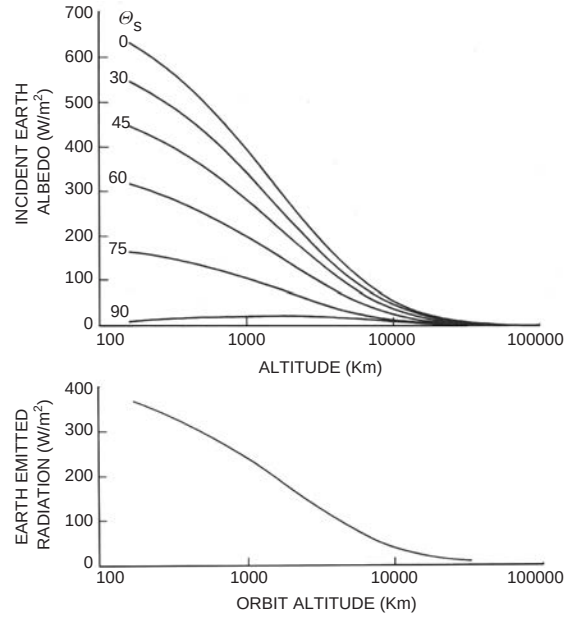


Figure 4.3.4: *Received albedo radiation (above) and infrared radiation (below) based on the cross-section of a spherical satellite as a function of orbit height [4.3.5].*

on the size of the Sun-illuminated Earth surface, on the **solar zenith angle** (the angle between the center of the Earth–satellite and the center of the Earth–Sun vectors), and on the inclination of the orbit. Because of the thermal capacity of the external satellite materials, the albedo variations in most cases have only a small influence on the satellite temperatures. The effects will appear first at the outermost layer of the satellite insulation due to its small heat capacity, but they are not critical there. The received albedo radiation decreases rapidly with higher orbits and is negligible for geostationary satellites. This is illustrated in Figure 4.3.4 (upper part), where Θ_S designates the angle between the Earth–Sun and Earth–satellite vectors.

The radiation emitted from the Earth is within the **infrared wavelength region** and corresponds to a black body temperature of about 250 K. The average radiation intensity in the near-Earth orbit amounts to 230 W/m^2 and may vary between 150 and 350 W/m^2. As shown in Figure 4.3.4 (lower part), the influence also decreases here with orbit height. Some values are given in Table 4.3.1. The numbers in this table indicate that for an orbital inclination of 90°, as an example, the radiation originating from the Earth amounts in 97% of the cases to 250 W/m^2. The expression "percentile" is the probability that a given value for albedo or infrared radiation will not be exceeded. All values in Table 4.3.1 are corrected for a 0° zenith angle.

Table 4.3.1: *Albedo and infrared radiation (percentile values).*

Orbital inclination	3%	50%	97%
Albedo radiation			
30°	0.14	0.18	0.22
60°	0.17	0.23	0.32
90°	0.18	0.25	0.34
Infrared radiation			
30°	227	246	265
60°	211	233	255
90°	205	227	250

The occurrence of deviations from average values of albedo and infrared radiation is described in [4.3.2] and in more detail in [4.3.4].

4.3.2.2 Radiative Heat Transfer

Heat radiation is the transport of energy by electromagnetic waves in the wavelength range between 100 nm and 100 μm. A body emits energy into the half space above its surface which is proportional to its surface F and to the fourth power of its temperature T. The radiation intensity depends on the properties of the radiating surface and is expressed by its **ability to radiate** or by the **emissivity** of the surface.

The emitted energy is expressed in

$$P_\varepsilon = \sigma\, \varepsilon\, T^4\ F \tag{4.3.2}$$

with $\sigma = 5.67 \cdot 10^{-8}\,\mathrm{W/(m^2\,K^4)}$ the Stefan–Boltzmann constant.

When a body is exposed to radiation, a part will be absorbed, a part reflected and a part transmitted through the body. The impinging radiation follows the expression

$$\alpha + \beta + \gamma = 1 \tag{4.3.3}$$

where:
α = solar absorption,
β = reflection,
γ = transmission.

For satellites the transmission is less important, since all thermally relevant materials are not translucent.

For the absorbed energy the following applies:

$$P_\alpha = \alpha \cdot S \cdot F_\mathrm{p} \tag{4.3.4}$$

where:
S = radiation intensity or solar flux density,
F_p = projected area.

The **Kirchhoff radiation law** postulates that an arbitrary surface absorbs energy to the extent that it emits energy, as long as the spectral distribution of the absorbed and emitted radiation correspond to each other. Therefore

$$\alpha_\lambda = \varepsilon_\lambda \tag{4.3.5}$$

In practice the symbols α and ε do have different meanings.

The term **(solar) absorptance** defines the capability of a surface to absorb radiation in the solar spectrum. Solar energy, compared to the radiation of a black body at 5760 K, covers a wider spectral band (Figure 4.3.5). Of the total radiation energy, 97% is located in the wavelength range between $\lambda = 0.2$ and 2.8 µm, with a maximum at about 0.48 µm.

The average solar absorptance is defined for this wavelength region. The spectral radiation flux density E (irradiance) can be defined with the **Planck radiation law**. Integration over all wavelengths and the radiation density over the half space gives the total energy or the radiation flux density in $\mathrm{W/m^2}$ according to the Stefan–Boltzmann law.

The **(thermal) emissivity** depends on the temperature of the radiating body. However, for the temperature range applicable to a spacecraft, emissivity values of about 300 K can be applied. Consequently, and according to **Wien's displacement law**, the radiation is located within the infrared band with a maximum at about 8 µm. Within the infrared band ($\lambda = 1$ to 100 µm), 95% of the total radiation energy is within the bandwidth of $\lambda = 5$ to 50 µm, which defines the thermal emissivity ε.

In space technology different **surface coatings** are applied to achieve a desired influence on the thermal behavior. The solar absorption relative to the wavelength of some common surfaces is given in Figure 4.3.6.

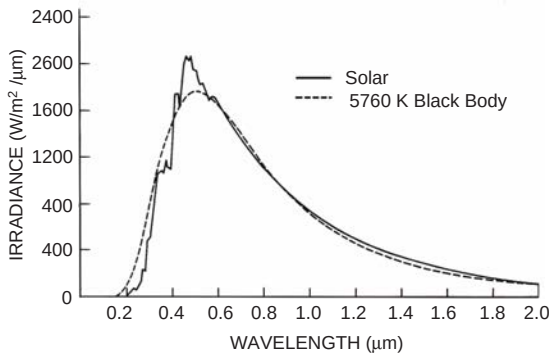


Figure 4.3.5: *Solar spectrum and spectral radiation flux density of a black body at 5760 k of thermal control [4.3.5].*

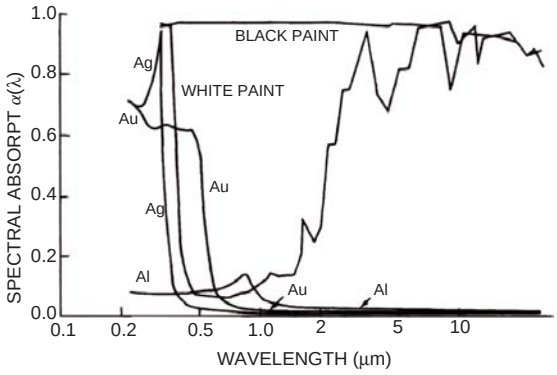


Figure 4.3.6: *Solar absorption of some common spacecraft surfaces [4.3.5].*

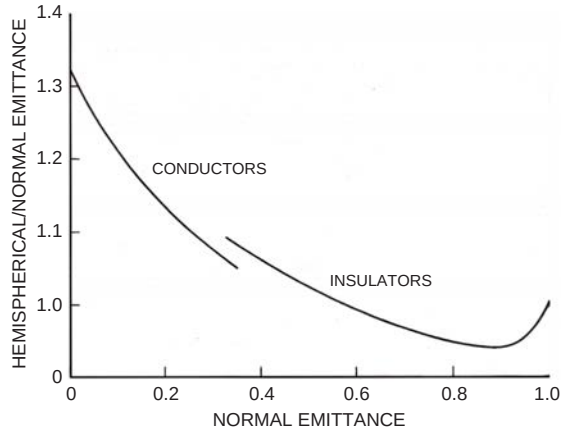


Figure 4.3.7: *Relationship between hemispherical and normal emissivity [4.3.5].*

Table 4.3.2: *Solar absorptance and infrared emissivity of materials mentioned in Figure 4.3.6, according to [4.3.5].*

Surface Coating	Solar absorption (α)	Emissivity under perpendicular radiation at 300K (ε_N)	Hemispherical emissivity at 300K (ε_H)	α/ε_H
White Paint	0.21	0.91	0.86	0.244
Black Paint	0.97	0.92	0.87	1.11
Vapor-Deposited Aluminum	0.08	0.018	0.024	3.23
Gold (Au)	0.19	0.015	0.020	9.5
Silver (Ag)	0.05	0.010	0.013	3.9

Figure 4.3.6 reveals that the absorptance of the special **white paint** applied in spacecraft thermal control is very low in the highest solar energy area (0.48 µm), but high in the infrared band. According to the relationship expressed in Equation 4.3.5, such a surface absorbs a small amount of solar energy but emits a high amount of energy in the infrared band. The surface will cool down also under direct solar radiation and is therefore suitable as a spacecraft heat sink (radiator). **Black paint** exhibits in the total spectral band high values of both α and ε. Such a surface is therefore used inside the spacecraft to ensure good heat exchange by radiation. Silver coatings exhibit small absorption coefficients in the range > 0.4 µm; gold (Au), however, only has this behavior starting from 0.7 µm.

The α and ε values for materials mentioned in Figure 4.3.6 are summarized in Table 4.3.2. The emissivity under perpendicular radiation can be directly determined by measuring the reflectivity. Hemispherical values, which are usually applied in thermal control, can be graphically determined from Figure 4.3.7 [4.3.5]. In many cases manufacturers of specific surface coatings specify the hemispherical emissivities which should preferably be used.

The relationship between absorptance and emissivity for spacecraft surfaces can be explained using

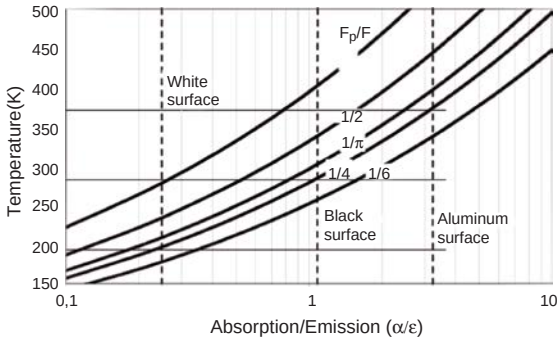


Figure 4.3.8: *Temperature versus ratios α/ε and F_p/F.*

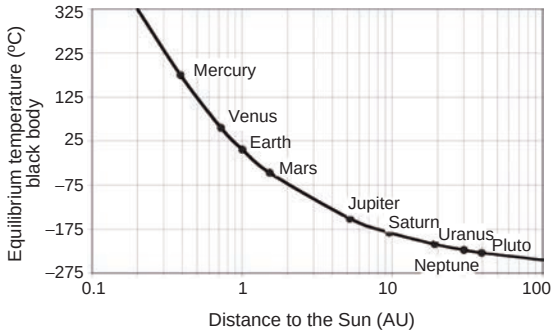


Figure 4.3.9: *Temperature of a black sphere in relation to the distance to the Sun.*

simple examples. For a body which is only exposed to solar radiation the **energy or heat balance** reduces to

$$P_\alpha = P_\varepsilon \tag{4.3.6}$$

Substituting Equations 4.3.2 and 4.3.4 into Equation 4.3.6 gives after some rearranging

$$T = \sqrt[4]{\frac{S}{\sigma}\frac{F_p}{F}\frac{\alpha}{\varepsilon}} \tag{4.3.7}$$

Equation 4.3.7 is plotted versus the relationships α/ε and F_p/F in Figure 4.3.8.
The symbols mean:
$F_p/F = 1$ Plane surface (back side insulated)
$F_p/F = 1/\pi$ Cylinder (insulated end surfaces)
$F_p/F = 1/4$ Sphere.

Figure 4.3.8 shows that a spherical body ($F_p/F = 1/4$) with a white, black or aluminum surface reaches temperatures of 195, 285 and 370 K, respectively. The property of white paint to achieve the lowest temperature is especially evident. On such a body considerable additional heat (dissipation in the case of a spacecraft) can be imposed in order to bring the temperature into the area of operating electronics.

The conditions for bodies at different distances from the Sun can be calculated with Equation 4.3.7 by introducing **solar intensities** corresponding to the distance to the Sun. The radiation intensity is inversely proportional to the square of the distance ($\sim 1/r^2$). This relationship is shown in Figure 4.3.9, where the temperature of a black sphere is plotted versus the distance to the Sun.

For calculating equipment temperatures, which are in radiation exchange with the environment, it may be useful to define an **effective radiation background temperature**. Such a background temperature includes the natural radiation (Sun, albedo, infrared) as well as the radiation exchange with other objects. According to Equation 4.3.6, the energy balance can be written as follows:

$$P_{S,i} + P_{A,i} + P_{IR,i} = \sigma \sum_{j=1}^{n} r_{ij}\left(T_j^4 - T_i^4\right) + \sigma\,\varepsilon\,T_H^4\,F \tag{4.3.8}$$

where:
$P_{S,i}$ = absorbed solar radiation on body i,
$P_{A,i}$ = absorbed albedo radiation on body i,
$P_{IR,i}$ = absorbed infrared radiation on body i,
T_i = temperature of body i,
T_j = temperature of body j,
r_{ij} = radiation coupling between bodies i and j,
F = surface of the background radiator.
Rearranging Equation 4.3.8 leads to

$$T_H = \sqrt[4]{\frac{P_{S,i}}{\sigma\,\varepsilon\,F} + \frac{P_{A,i}}{\sigma\,\varepsilon\,F} + \frac{P_{IR,i}}{\sigma\,\varepsilon\,F} - \frac{\sum_{j=1}^{n} r_{ij}\,(T_j^4 - T_i^4)}{\varepsilon\,F}} \tag{4.3.9}$$

Radiation coupling is described in more detail in Section 4.3.3.2. The background temperature for thermal radiation, which is according to Equation 4.3.9 only valid for equilibrium conditions, serves as a simple method to conduct **parameter studies** without

calculations and to define boundary conditions for thermal tests in vacuum chambers.

4.3.2.3 Heat Conduction

Heat conduction is defined as the transport of heat between two locations of a solid body due to a temperature gradient. The basic process is the transport of kinetic energy between molecules and consequently heat conduction takes place also within liquids and gases. Due to the necessary presence of molecules, heat conduction is not possible in a vacuum.

Heat transport can be expressed by

$$\dot{Q} = \frac{\mathrm{d}Q}{\mathrm{d}t} = \lambda F \frac{\Delta T}{d} \tag{4.3.10}$$

where:
$\mathrm{d}Q/\mathrm{d}t$ = energy received/released by heat conduction or heat flux per time unit,
λ = heat conductivity,
F = cross-section of the heat path perpendicular to the line connecting locations 1 and 2 of the solid body,
ΔT = temperature gradient between locations 1 and 2 of the solid body,
d = distance between locations 1 and 2 or thickness of the material.

Differentiating Equation 4.3.10 with respect to the area gives the **heat flux density** or the heat flux per unit of time and area:

$$\dot{q} = \frac{\mathrm{d}\dot{Q}}{\mathrm{d}F} = \lambda \frac{\Delta T}{d} \tag{4.3.11}$$

The **thermal resistance** R_{th} [K/W], which is also a useful material property, is defined as

$$R_{\mathrm{th}} = \frac{d}{\lambda F} \tag{4.3.12}$$

which by insertion in Equation 4.3.10 becomes

$$\dot{Q} = \frac{\Delta T}{R_{\mathrm{th}}} \tag{4.3.13}$$

A **heat flux** $\mathrm{d}Q/\mathrm{d}t$ behaves similarly to an **electric current** I in Ohm's law, where the temperature difference ΔT corresponds to the electrical voltage U and the thermal resistance R_{th} to the electrical resistance R_{el}. Therefore, R_{th} corresponds to $\Delta T = R_{\mathrm{th}} \cdot \mathrm{d}Q/\mathrm{d}t$. Consequently, thermal–mathematical models can be constructed similarly to electric circuits. As an example, the heat flux through three materials positioned in series can be determined by

$$\dot{Q} = \frac{\Delta T}{R_{\mathrm{th},1} + R_{\mathrm{th},2} + R_{\mathrm{th},3}} \tag{4.3.14}$$

4.3.2.4 Thermal Contact Resistance

Equation 4.3.12 is valid for a perfect contact between the materials concerned. In real space systems roughness and unevenness between different elements need to be considered. Consequently, Equation 4.3.12 transforms into

$$\dot{Q} = \frac{\Delta T}{R_1 + R_{\mathrm{ct},1} + R_2 + R_{\mathrm{ct},2} + R_3} \tag{4.3.15}$$

where R_{ct1}, R_{ct2} is the thermal contact resistance.

The thermal **contact resistance** is defined as the relationship between the temperature difference and the heat to be transferred, or

$$R_{\mathrm{ct}} = \frac{\Delta T}{\dot{q} F} \tag{4.3.16}$$

with F the visible contact surface.

In the literature the **thermal contact conductance** is often defined and expressed as

$$h_{\mathrm{ct}} = \frac{1}{R_{\mathrm{ct}} F} \tag{4.3.17}$$

The **heat transport processes** taking place between two contacting bodies are very complex and depend amongst other factors on the following parameters:

- Roughness of the contacting surfaces in the microdomain and unevenness in the macrodomain.
- Gap thickness.
- Existence of a fill medium in the gap (vacuum, thermal grease, foil, etc.).
- Heat conductivity of contacting materials and the fill medium.
- Hardness of the materials (which defines the pressure for plastically deforming the spikes of the rough surfaces).

Table 4.3.3: *Heat conduction value for bolted connections and a graphite fiber foil as interface filler [4.3.7].*

Fixation torque[Nm]	Temperature [°C]	Thermal contact conductance [W/m² K]
1.13	–34	284
1.13	+71	369
2.26	–34	329
2.26	+71	397

- Module of elasticity and thickness of the materials (influences the elastic deformation for decreasing the unevenness).
- Temperature in the contact area (influences the materials properties).

Due to this complexity a standard theory to determine heat transfer is not available. Different theoretical approaches are described in [4.3.6]. The development of a **thermal control system** is commonly based on test results and engineering practice.

Due to the **roughness** of the contacting bodies, both surfaces exhibit a fissured structure in the microdomain and consequently only microspikes on the surfaces are in contact. Statistically the number of these contacts increases with a decrease in roughness. By increasing the contact pressure some spikes will be plastically deformed, leading to a larger contact surface. In areas without any contact, heat transfer may be accomplished by radiation or by conduction through an enclosed gas (in terrestrial applications, mostly air). Due to the small thickness of the gas layer, the heat transport may be significant despite the low thermal conductivity of the gas (the small dimensions are also the reason why heat transport by convection does not take place).

In space applications the contribution of a gas to heat conduction does not exist. However, in order to utilize the portion of a surface which does not have a direct contact for heat transport, thin foil-like **filler materials** are placed in the contact area. These interface fillers are selected for high heat conductivity and a suitable hardness, which enables the fissured surface structure to work into the filling material, leading to good contact across the entire surface. The attainable

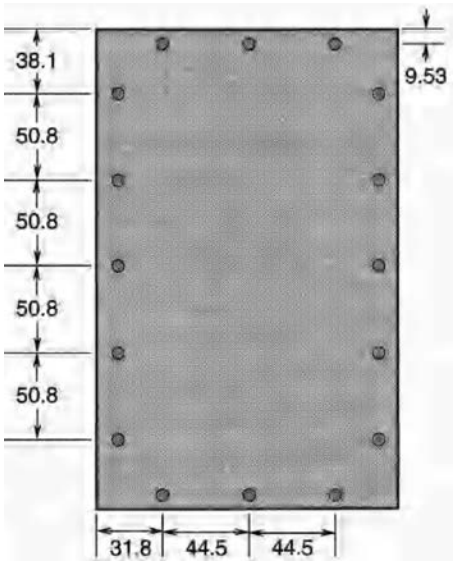


Figure 4.3.10: *SIGRAFLEX foil as interface filler for the contact areas [4.3.6].*

heat transport depends in addition on the local contact pressure of the two surfaces. Electronic housings are mostly bolted along the circumference, producing the maximum heat fluxes there. This is favorable for housings which use the side walls to transport electrical waste energy to the base plate.

Table 4.3.3 shows test results from a 152 × 279 mm² plate connection which is pressed together at the edges with 16 stainless steel bolts. The tests were performed with two Al 6063-T6 plates of 7.94 mm thickness each. A graphite-fiber foil was used as interface filler. It can be clearly seen that both a higher fastening torque as well as higher test temperatures increase the heat transfer coefficient or the heat conduction value.

In the European space industry the graphite foil SIGRAFLEX produced by SGL CARBON AG is mostly used. The material is also mentioned on the ESA Preferred Materials List (Figure 4.3.10).

The foil is 0.2 mm thick with a heat conduction of 150 W/(m K) parallel to and 4–6 W/(m K) perpendicular to the surface. With this interface filler contact conduction values of 500 W/(m²K) can be obtained for regular electronic housings. For smaller surfaces (20 × 20 mm²) this value increases to about 1000 W/(m²K).

4.3.2.5 Mathematical Thermal Model

In the previous sections the temperature distribution was assumed to be time invariant ($\partial T/\partial t = 0$). A spacecraft is a complex structure with a multitude of different parts and components. Since an analytical solution for the temperature distribution of an inhomogeneous body does not exist, the problem is treated with a thermal model consisting of the **discrete mass and surface elements** (called **nodes**) of the space vehicle. Such a mathematical node model is a baseline for determining the time-variant temperature distribution of a spacecraft.

The **thermal energy** Q is proportional to the temperature T and expressed by the relationship

$$Q = m\, c\, T \tag{4.3.18}$$

where:
c = specific heat capacity,
m = mass of the node or the heat sink.
Differentiating equation (4.3.18) with respect to time leads to

$$\frac{\partial Q}{\partial t} = m\, c \frac{\partial T}{\partial t} \tag{4.3.19}$$

Balancing between received and released energy for a node i, the relationship (4.3.19) expands to

$$m_i\, c_i \frac{\partial T_i}{\partial t} = (\alpha_i\, \dot{q}_{A,i}\, e_{A,i} + \varepsilon_i\, \dot{q}_{IR,i}\, e_{IR,i} + \alpha_i\, S\, e_{S,i}) \cdot F_i + \sum \Delta \dot{Q} + \dot{Q}_i \tag{4.3.20}$$

where:
$\dot{q}_A$ = albedo flux density,
$\dot{q}_{IR}$ = flux density of the Earth's eigenradiation,
S = solar flux density,
F = radiating surface,
$e_{A,IR,S}$ = view or configuration factors,
dQ/dt = heat dissipation.

The term $\sum\Delta(dQ/dt)$ is the energy, subdivided into heat conduction and radiation, of a node i, which is either received from or emitted to surrounding nodes. The thermal mathematical model therefore consists of a system of nonlinear, ordinary first-order differential equations of the following form:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(t, \boldsymbol{x}) \quad \boldsymbol{x} = (x_1, \ldots, x_n)^T, \tag{4.3.21}$$

$f=(f_1,\ldots,f_n)^T \in \Re^n$

for n unknown **node temperatures**, which are solved with numerical calculation methods. Equation 4.3.20 is valid for **edge nodes**, that is for nodes which are exposed to the external radiation field. For all inner nodes of the spacecraft the following applies:

$$s = \dot{q}_A = \dot{q}_{IR} = 0 \tag{4.3.22}$$

and Equation 4.3.20 simplifies to

$$m_i\, c_i \frac{\partial T_i}{\partial t} = \sum \Delta \dot{Q} + \dot{Q}_i \tag{4.3.23}$$

In addition, for inner, structural "nodes" the dissipation term does not apply ($dQ/dt = 0$).

For the preliminary design during a definition phase a low number of nodes (< 50) is normally sufficient. In follow-up development phases the **number of nodes** is considerably increased (100 to 500) in order to accurately investigate and evaluate the different heat fluxes. The finally selected number of nodes depends on the complexity of the structure and on the experience of the thermal engineer. Basically one can state: "as many nodes as necessary and as few as possible."

The real effort is connected with the determination of heat conductivities and radiation factors, which are used to interlink the nodes and the thermal relationships. This issue and other aspects in the development of a thermal system are detailed in the next section. The determination of geometrical view or configuration factors is not explained here, but can be found in the literature, for example [4.3.8].

4.3.3 Development of the Thermal System

4.3.3.1 Overview

According to the progress of a space project and parallel to the development of the entire space vehicle

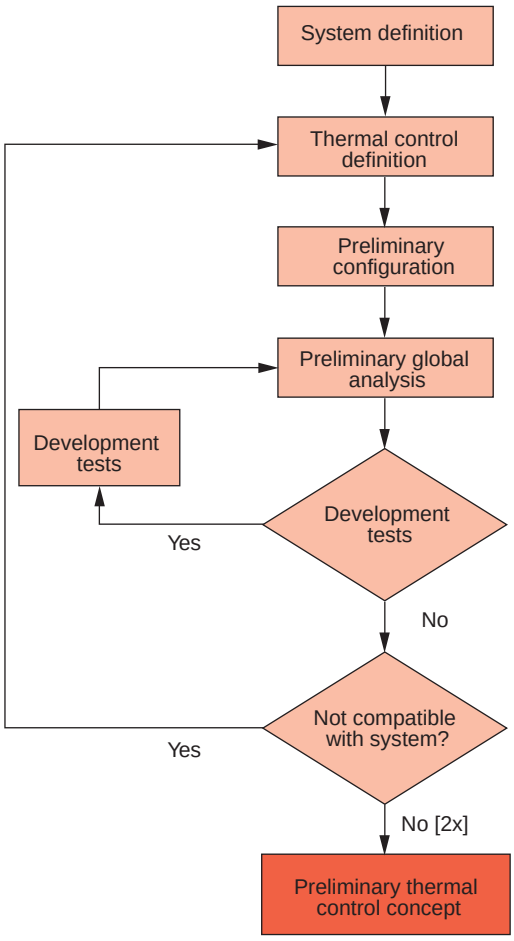


Figure 4.3.11: *Development steps for a preliminary thermal system (according to [4.3.9]).*

system, the development of a thermal system takes place in three steps (Figure 4.3.11 and Figure 4.3.12):

- Preliminary design presented in the context of the preliminary design review (PDR) of the system.
- Final design presented in the context of the critical design review (CDR) of the system.
- Verification of the thermal design by testing in a vacuum chamber in a simulated space environment.

First analysis cases are performed based on preliminary **system requirements** and **subsystem specifications**, which leads to the definition of a preliminary

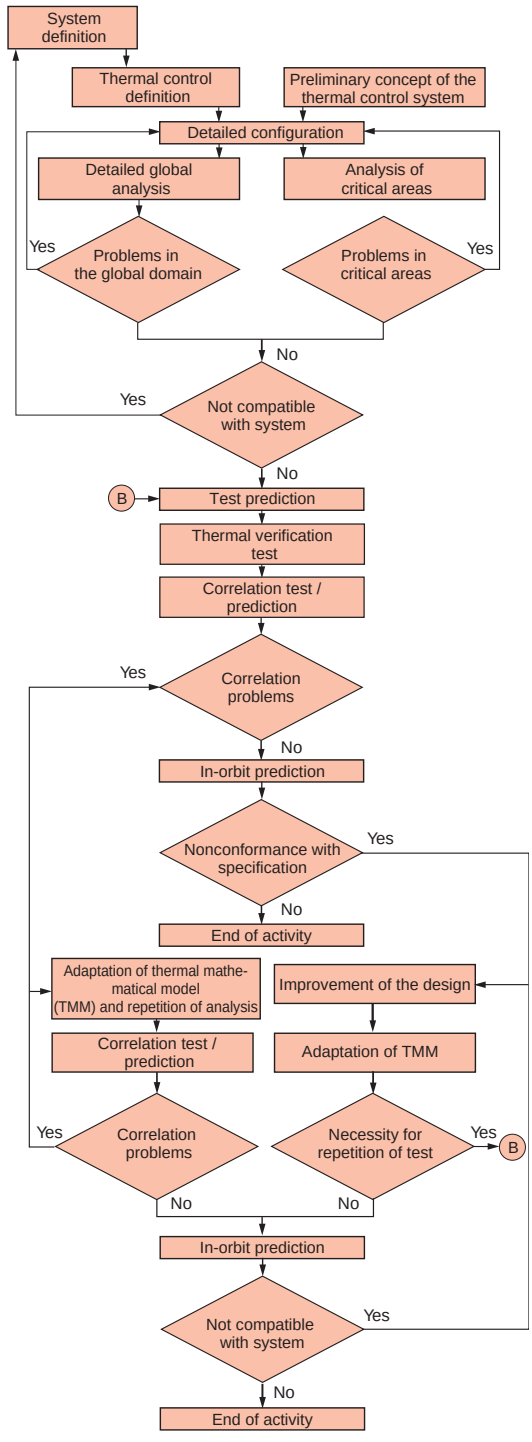


Figure 4.3.12: *Development steps for a final thermal system (according to [4.3.9]).*

thermal configuration. Several analytical iterations are generally necessary, since system requirements are subject to evolution at the beginning of a project. The preliminary design already defines the size and position of radiators, determines the layout and electrical power of heaters, and describes thermal requirements for electronic equipment (thermo-optical properties of surfaces, contact surfaces, distribution of heat dissipation, etc.).

Detailed analysis of the thermal system generally starts with updated requirements derived from the preliminary design review of the system. The **definition of the thermal system** is continued based on a detailed analytical overall model of the spacecraft, with critical areas being treated in parallel with smaller, but more detailed models. The **final design** is presented at the CDR of the system and is then released for the manufacturing phase.

The verification of all requirements is finally accomplished with a **thermal balance test**, which is generally performed after the thermal behavior has been determined by complex radiation and conduction processes. Generally the entire satellite is exposed to simulated space environments (vacuum, low temperature). Test aims include:

- Collection of data for **verification** of the mathematical spacecraft model as part of the thermal system qualification.
- Demonstration of the suitability of the thermal design.
- Verification that the thermal system meets the requirements.
- Information about the sensitivity of the thermal design to **parameter variation**, in particular material properties and heat dissipation of the electrical equipment.

In order to meet the test aim, the test object is put into thermal equilibrium and the resulting temperatures are then compared to analytical test predictions. As a rule, extreme cold and warm orbit cases are simulated. For test objects with complex geometries, noncontinuous surfaces or, in cases where the absorbed solar flux can be determined only with uncertainty, simulation of the solar radiation is recommended in addition. (The large European test facilities are equipped with the relevant installations.)

4.3.3.2 Analytical Methods

As mentioned in Section 4.3.2.3, thermal problems can be solved in **analogy to electrical laws** and networks (Ohm's law, Kirchhoff's law). These networks are described as partial differential equations for which there are different solution methods. Discrete approximation using a network has some advantages for numeric as well as experimental solutions. A **network model** follows for steady-state cases a system of algebraic equations with a finite number of variables, which may be solved with standard techniques. Transient cases result in first-order differential equations, which are numerically integrated.

In particular in space engineering, **finite difference and finite element methods** are being applied. For both processes the systems to be treated are divided into networks of subelements, to which different properties are assigned. The reduction of a simple body to three subelements for evaluation with the finite difference method is shown in Figure 4.3.13.

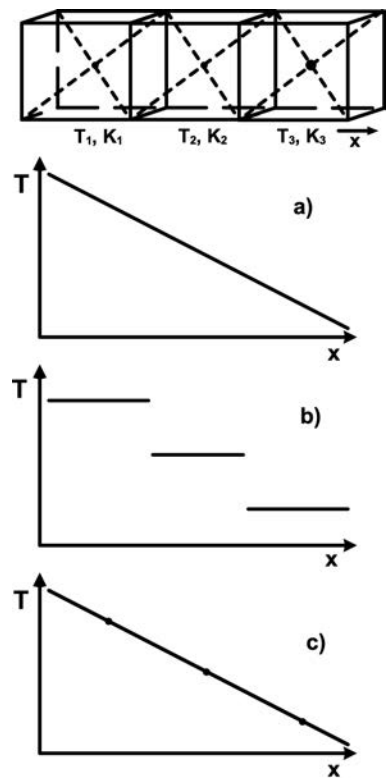


Figure 4.3.13: *Reduction of a body into subelements.*

Each element exhibits an average temperature T and a heat capacity K.

The latter value is calculated from material properties. By using this method the properties of the subelements, including temperature, are assumed to be located in the center of the element. Such an approach is called the **method of concentrated parameters** or the **lumped parameter method**. The interpolation of the point results (Figure 4.3.13c) leads to an approximation of the ideal distribution (Figure 4.3.13a). The transition to a step-like representation (Figure 4.3.13b) is avoided. The method is practically a finite difference approximation of the higher order differential equation 4.3.20. The described subelements exhibit a finite thermal mass and capacity and are called diffuse nodes. In thermal software packages (e.g., ESATAN) three node types are defined:

- Diffusion nodes with finite thermal mass
- Arithmetic nodes without a thermal mass
- Boundary nodes with indefinite thermal mass.

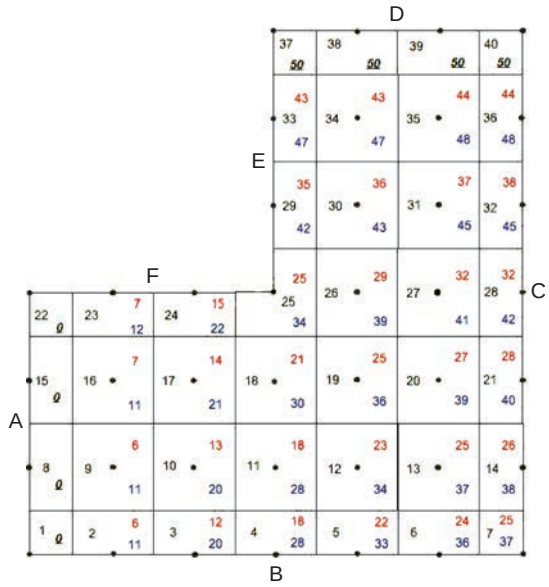


Figure 4.3.14: *Example of node distribution on a plate.*

The **diffusion node** represents the result of a thermal analysis by its temperature and is based on a heat flux entering or leaving the node.

Arithmetic nodes are not really meaningful in the strict physical sense, but helpful when presenting surfaces in a model or describing components with very small masses (bolts, foils, insulation with small mass, or small fluid volumes in tubes). Arithmetic nodes, which constitute a small subset of nodes relative to the total number, adopt the temperature of the immediate environment.

Boundary nodes define lines, surfaces and points of constant temperature in a model, and this independently of the receiving or emitting heat flux. They represent, for example, the background temperature of a satellite (space as heat sink, planetary temperature) or in relation to other nodes of very high mass (liquid in a large tank).

The detailed reduction of a body to nodes and volume elements is performed to optimally represent the expected results, including the minimization of calculation time (computer runtime).

Generally the following criteria are valid:

- Identification of locations for which temperatures are to be calculated.
- Required verification of temperature gradients.
- Taking into account the geometrical complexity (possible simplifications).
- Physical plausibility.
- Accuracy of results vs. computing time.

Since all these factors depend on each other, problem-related approaches are based on practical experience and engineering judgment. Volume elements are in practical cases simple geometrical shapes (cube, rectangle, etc.) with properties which can be easily calculated (surfaces, volumes, mass, capacity). Complex structures can be approximated by simple shapes, where the method of approximation depends on the required accuracy of the results. Again, this is mostly based on engineering practice.

The process of the **lumped parameter method** is explained below using a simple example from [4.3.10].

A metallic plate corresponding to Figure 4.3.14 is assumed. The border lines A and D have fixed temperatures of 50 and 0 °C, respectively, while the border lines B, C, E and F are insulated. The plate as shown in Figure 4.3.14 has thickness d, thermal conductivity λ, specific heat capacity c and density ρ. The steady-state temperature distribution and the transient behavior for a start temperature of 0 °C at time $t = 0$ are to be

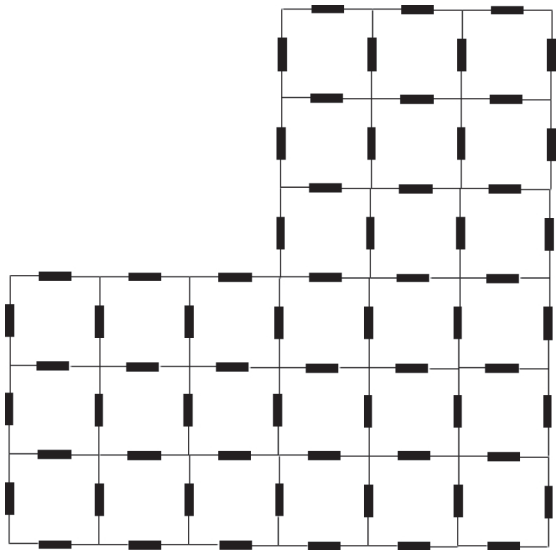


Figure 4.3.15: *Conductors of node model from Figure 4.3.14.*

determined. First the plate is divided into square elements. According to the selected geometry, the border and edge elements have half, quarter and three-quarter square shapes. Nodes are separated by a distance s = 50 mm in the center of the elements.

Nodes are also placed on the border of the plate, in order to define fixed temperatures as a boundary condition. Based on this definition (node distance equal to element edge length) the heat conductivities and capacities are calculated as follows.
Heat conduction in the center of the plate:

$$L_m = \frac{\lambda\, d\, s}{s} = \lambda\, d \tag{4.3.24}$$

Heat conduction at the plate edges:

$$L_r = \frac{\lambda\, d \frac{s}{2}}{s} = \frac{\lambda\, d}{2} \tag{4.3.25}$$

The heat capacities, which are needed for transient calculations, are derived from material properties: for center nodes,

$$K_m = d\, s\, s\, \rho\, c = c\, \rho\, s^2\, d \tag{4.3.26}$$

for edge nodes (half of element surface),

$$K_r = d\, s \frac{s}{2} \rho\, c = \frac{c\, \rho\, s^2\, d}{2} \tag{4.3.27}$$

for corner node 25 (one quarter of element surface),

$$K_e = d \frac{s}{2} \frac{s}{2} \rho\, c = \frac{c\, \rho\, s^2\, d}{4} \tag{4.3.28}$$

For the corner node (three-quarters of element surface),

$$K_{el} = \frac{3\, c\, \rho\, s^2\, d}{4} \tag{4.3.29}$$

After assigning node numbers corresponding to Figure 4.3.14 a discretization of the plate emerges with 40 nodes and 66 conductors (Figure 4.3.15).

This example has been analyzed with the software package ESATAN based on the following boundary conditions:

- Material: pure aluminum (thermal conductivity λ = 204 W/(m K), specific heat capacity c = 879 J/(kg K), density ρ = 2700 kg/m^3)
- Plate thickness d = 2 mm
- Edge nodes 1, 8, 15 and 22 with 0 °C as fixed temperature
- Edge nodes 37, 38, 39 and 40 with 50 °C as fixed temperature.

Results:

1. Steady-state temperature field (without radiation)
   Result: upper red numbers in Figure 4.3.14.
2. Steady-state temperature field with additional solar absorptance of α = 0.2 on one side and single-sided radiation loss of ε = 0.8
   Result: lower blue numbers in Figure 4.3.14.
3. Transient temperature behavior for the case that in the beginning all node temperatures are 0 °C
   Result: after 6 minutes the plate has steady- state behavior with temperatures as under 1.

Thermal software packages generally work with three different conductance values, as follows:

GL value = linear conductance according to Equation 4.3.11 between node i and j with temperatures T_i and T_j:

$$\Delta \dot{q}_{ij}^{c} = k_{ij}\left(T_i - T_j\right) \qquad (4.3.30)$$

with k_{ij} the heat conductivity between nodes i and j. Linear conductance values represent the heat transport between solid bodies and processes which can be linearly treated, such as convection, condensation and evaporation.

GR value = heat exchange by radiation according to (4.3.1), expressed as heat flux density:

$$\Delta \dot{q}_{ij}^{\mathrm{rad}} = \sigma\, h_{ij}\left(T_i^4 - T_j^4\right) \quad \text{with} \quad h_{ij} = \varepsilon_i\, \varepsilon_j\, e_{ij} \qquad (4.3.31)$$

h_{ij} is also the radiation shape factor; it contains the effective emissivity as well as the view factor between nodes i and j. The relevant user manuals for commercial computer codes (e.g., ESATAN) contain details on how to create from this information an initial data set for the code, how to specify the solution method, and how to present the results.

GF value = the heat flux in or out of a node due to fluid flow in a tube. It is assumed that the fluid enters node i with temperature T_u of the upstream node u and leaves the node with its own temperature T_i. Therefore the heat flow can be written as

$$\Delta \dot{q}_{i}^{fl} = w\, c\left(T_u - T_i\right) = l\left(T_u - T_j\right) \qquad (4.3.32)$$

where:
w = mass flow density,
l = GF value.

Temperature calculations are also possible using the **finite element method** (FEM), which, however, is most often used for structural analysis. Classical applications are found in the borderland between thermal and structural problem areas, for example to analyze thermomechanical stresses and thermally caused structural distortions. By using the FEM a **finely woven mesh of subelements** is placed over the analyzed structure. One-, two- or three-dimensional elements can be used, which may adopt different geometries, as shown in Figure 4.3.16. Each element has element nodes at the corners (not in the center as for the finite difference method). Parameters (e.g., temperatures) are generally assigned to and calculated for the edge nodes and interpolated across the element, if they

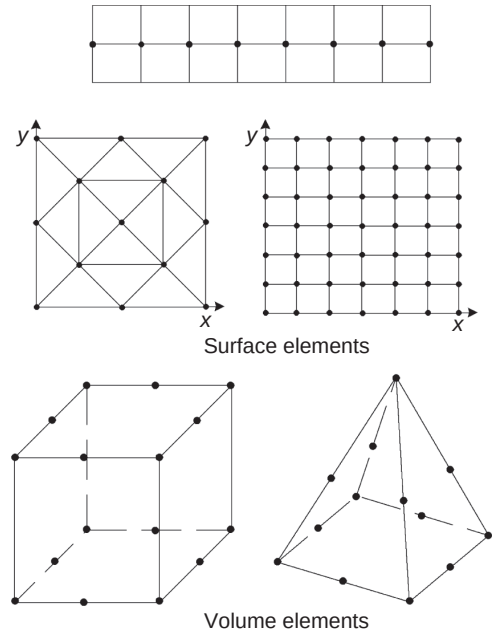


Figure 4.3.16: *Node and element definition for the FE method.*

differ. Consequently, properties of an element are not constant (in contrast to the isothermal nodes of the finite difference method).

The FEM calculates an explicit expression for the temperature, based on known relationships which satisfy the governing differential equations and the boundary conditions of the element. The method is very versatile and able to solve steady-state, dynamic, thermal and electromagnetic problems, as well as nonlinear problems (distortions, nonlinear material properties, contact problems).

There are reasons why the FEM is not being used all that often in thermal analysis. Since the accuracy of an FE analysis increases with the number of elements, namely the detail of the element network, an FE model generally uses some hundreds to thousands of nodes in order to obtain adequate results. This is a disadvantage for calculating thermal problems, since, as shown above, analyzed temperatures are assigned to each node. Such a high number of calculated temperatures is not necessary for a thermal analysis.

Diffuse nodes (nodes with mass) and edge nodes are also used in the FEM; however, arithmetic nodes (those without mass) cannot be included. In addition,

the FE methodology cannot assign a single node to a larger body (e.g., an electronic unit), as is done in the finite difference process. As a result, the FE model for typical thermal analysis cases is always larger than necessary. In addition, curved surfaces like cones and cylinders require many more finite surface elements to describe the configuration than are used for the finite difference method. For the latter method a 360° node is often sufficient, which cannot be introduced into the FEM. While for radiative heat exchange according to the Monte Carlo method the true representation of a curved surface is accepted, these surfaces in the FEM are represented by flat elements or they are approximated by a polynomial adaptation. As a consequence, the accuracy of the radiation view factor calculations may be decreased, which would again require compensation with an increased number of elements. The combination of an increased number of surface elements and resulting view factors may in addition increase the complexity and duration of the thermal analysis.

4.3.3.3 Verification by Testing

After the analytical design phase and after hardware manufacturing, a test phase follows with the aim of demonstrating that the hardware fulfills specified requirements. In general there is a distinction made between three test categories.

Development tests are conducted very early in the program to determine the usefulness of new technical concepts and to reduce risks before the manufacture of flight hardware commences. Experience obtained during development tests flows into procedures for qualification and acceptance tests for flight hardware designs.

Qualification tests are formal demonstrations that the design and manufacturing process produces correct flight hardware which meets the requirements. These tests also validate acceptance test procedures, including test methodology, instrumentation and necessary software. In order to meet these aims, the test conditions are more stringent than the expected flight conditions. Qualification hardware is therefore not flown.

Acceptance tests are again formal tests which demonstrate the adequacy of the supplied hardware for application as flight hardware. The test demonstrates that the specified requirements have been met and that no hidden failures are present due to manufacturing error or material selection.

In the mentioned test categories different thermal tests are performed, as follows.

During the **thermal cycle test** the test object is alternately exposed to cold and warm temperatures. On an equipment level these tests are mostly conducted in air and the subsequent convection helps to accelerate the change between the temperature extremes. Thermal stress due to thermal cycling uncovers hidden flaws in the material and any mounting errors that may have occurred during integration of the test object. As a second goal functional tests are performed which need to be successful at the extreme temperatures.

Also during **thermal vacuum testing** the test object undergoes temperature cycles. Since the natural convection is eliminated, this test represents real space conditions and realistic functional tests are possible. Since the change between temperature extremes is slower here (the lack of convection), the thermal stress behavior is less important.

A **thermal balance test** is generally performed on the subsystem or system level and is mostly part of the above described thermal vacuum test. This is the most important test for a thermal subsystem and has two important aims:

- Verification of the thermal design
- Correlation with the mathematical simulation model.

Flight conditions which are decisive for the design of the thermal system are derived for this test and test results are compared to the analytical prediction which has been determined with the thermal mathematical model. After successful correlation the mathematical model is verified and **temperature predictions** for the flight phase can be made. Test cases mostly include warm and cold flight conditions in order to cover all aspects of the subsystem design, including the operational heater system, dimensioning of the radiators, and critical heat fluxes. The large test centers in Europe offer the possibility also to simulate solar radiation, which is used for complex spacecraft configurations.

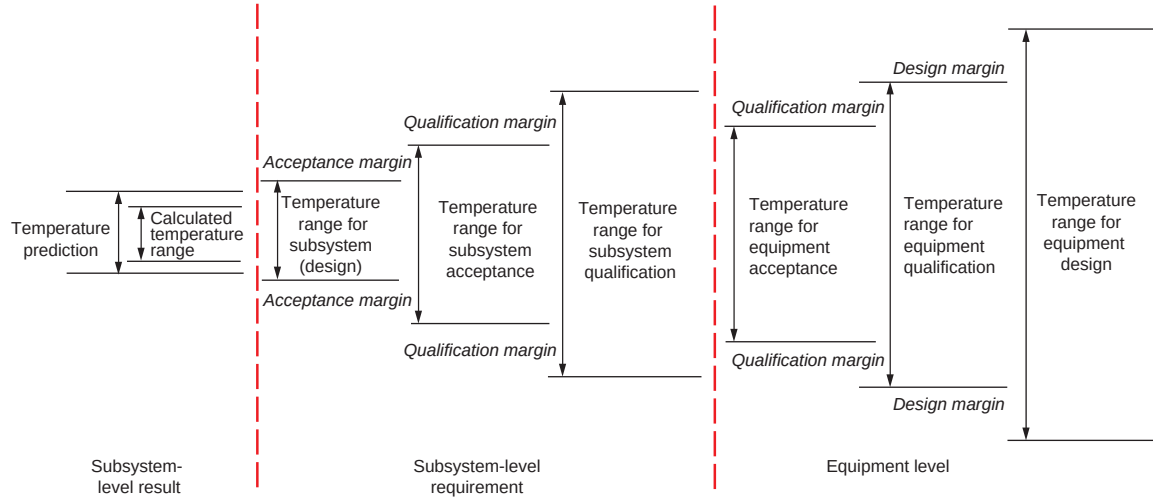


Figure 4.3.17: *Temperature definitions for the equipment and system levels (according to [4.3.9]).*

In recent times solar simulation has been selected to a lesser extent due to the high test costs. This trend is supported by the high quality and reliability of today's computer and simulation programs, which are able to predict the effects of solar radiation with high accuracy. The selection of test temperatures is based on the specified ranges for operational, nonoperational and switch-on temperatures for the equipment. These ranges are defined in Figure 4.3.17.

4.3.4 Technical Solutions

4.3.4.1 Overview

To optimize the thermal control system a combination of techniques is employed which decrease or increase the conductive or radiative heat flux in selected areas. The most important methods are the following:

- Decrease the radiative heat flux by **multilayer insulation** (MLI).
- Increase or decrease the **radiative heat exchange** with **surfaces** of high or low coefficients of absorptance and emissivity (e.g., black or metallic surfaces).
- Optimize the **radiation behavior** based on surfaces with low absorptance in the solar spectrum and high emissivity in the infrared region.
- Manipulate the radiative heat exchange by selecting appropriate geometries and stabilization (e.g., Helios A and B).
- Actively manipulate the radiative heat transfer by using **louvers** which automatically open or close at a predetermined temperature. The same effect can be achieved with electrically actuated coating, which changes the coefficient of emissivity.
- Employ **materials** with high thermal conductivity (aluminum, high-conductivity carbon fiber) or with low thermal conductivity (plastics).
- Use **two-phase cooling loops** to efficiently transport heat between distant spots in a spacecraft or to isothermalize radiator areas (heat pipes, loop heat pipes).
- Use of **heating elements** to regulate and stabilize the temperature of specific parts.

In the following the most important methods are explained in more detail.

4.3.4.2 Thermal Insulation

Thermal insulation is applied to minimize heat fluxes between two temperature regimes. Either homogeneous material with low thermal conductivity is used, which minimizes heat conduction, or multilayer

isolation, which reduces heat exchange by radiation since each layer acts as a radiation barrier.

MLI, which is used quite often for the thermal control of spacecraft, reduces the heat flux from and to a component and minimizes temperature fluctuations under varying external radiation environments. Typical applications are:

- Reduction of heat flux into a cryogenic fluid tank to minimize evaporation losses.
- Reduction of heat losses of a component to reduce necessary heater power in order to keep the unit within the specified temperature limits.
- Insulation of a spacecraft to reduce heat losses during shadow phases and to minimize heat introduction during direct solar radiation. This considerably decreases internal temperature variations caused by externally varying environmental conditions.

In an isolated system where forces and moments must be transmitted from a location of high temperature to a location of low temperature, or vice versa, thermal insulation may also be accomplished through specific design solutions. In such unconventional techniques the insulation effect of spherical contact surfaces is used; for example, commercial ball bearings from steel or glass are forced as a "layer" between two or more metal plates or between co-cylindrical tubes [4.3.11]. Depending on selected design options, force pattern and material properties, different thermal resistivity values and heat fluxes can be obtained.

MLI consists of several layers of **plastic foil** (polyester or polyamide) separated by plastic nets to reduce heat conduction. In order to minimize the thermal radiation exchange the foils are coated on both sides with aluminum ($\varepsilon \leq 0.025$). The thickness of the foils is between 15 and 50 µm and of the coating about 25 nm. MLI exhibits its excellent insulation effect in vacuum, since heat conduction via gas is eliminated. The foils are therefore perforated in order not to damage the MLI during evacuation and to reduce the remaining residual pressure, which should be below 10^{-8} Pa.

The **emissivity** of an evacuated MLI with n foils which do not touch each other (i.e., not considering gas and solid body conduction) can be expressed theoretically by

$$\varepsilon = \frac{1}{\frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2} - 1}\left(\frac{1}{n+1}\right) \tag{4.3.33}$$

where ε_1 and ε_2 are the emissivities of the two foil sides. In practice these ideal conditions do not exist. Therefore an **effective emissivity** ε^* and an **effective thermal conductivity** k^* are defined, based on test results:

$$\varepsilon^* = \frac{\dot{Q}}{F\,\sigma\left(T_w^4 - T_k^4\right)} \tag{4.3.34}$$

and

$$k^* = \frac{d\,(\dot{Q}/F)}{T_w - T_k} \tag{4.4.35}$$

where:

$\dot{Q}$ = heat transport through the MLI,
F = surface of the MLI,
T_w, T_k = temperature of the cold and warm sides of the MLI,
d = thickness of the MLI.

Table 4.3.4 summarizes measured values of ε^* and k^* for MLI with different foil layers and temperature ranges. However, both theoretical as well as measured values, which are based on large, flat MLI blankets, are not sufficient to predict the actual effectiveness of MLI

Table 4.3.4: *Effective conductance and emissivity of 10- and 20-layer MLI [4.3.12].*

T^* (in °C)	10 layers		20 layers	
	k^* (W/(m² K)	ε^*	k^* (W/m² K)	ε^*
−85	$0.5 \cdot 10^{-2}$	$2.7 \cdot 10^{-3}$	$0.4 \cdot 10^{-2}$	$2.2 \cdot 10^{-3}$
−45	$1.4 \cdot 10^{-2}$	$3.6 \cdot 10^{-3}$	$1.0 \cdot 10^{-2}$	$2.5 \cdot 10^{-3}$
+10	$2.1 \cdot 10^{-2}$	$3.6 \cdot 10^{-3}$	$1.3 \cdot 10^{-2}$	$2.3 \cdot 10^{-3}$
+50	$2.9 \cdot 10^{-2}$	$3.8 \cdot 10^{-3}$	$2.1 \cdot 10^{-2}$	$2.6 \cdot 10^{-3}$
+71	$3.9 \cdot 10^{-2}$	$4.2 \cdot 10^{-3}$	$2.8 \cdot 10^{-2}$	$3.0 \cdot 10^{-3}$

T^* = average temperature of the MLI.

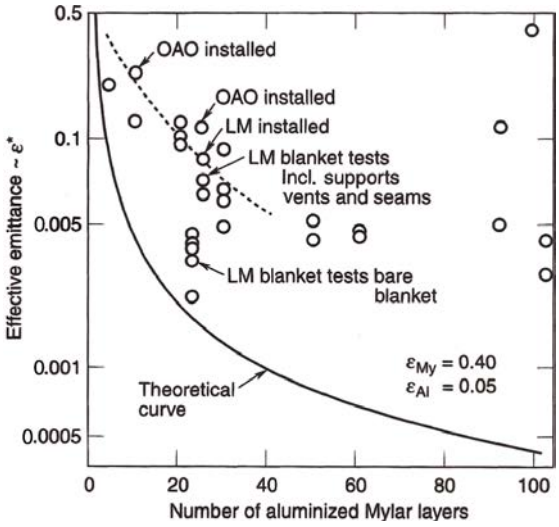


Figure 4.3.18: *Effective MLI emissivity in relation to the number of layers [4.3.6].*

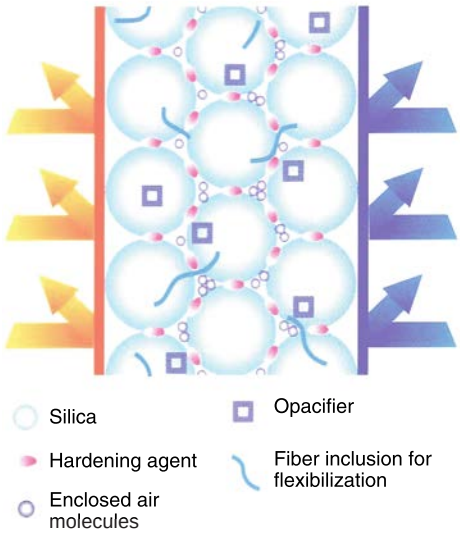


Figure 4.3.19: *Operational principle of an insulation based on silica [4.3.13].*

in an integrated configuration, since the following factors decrease the insulation property:

- **Residual gases** remaining between the foils, which may also be caused by continuous outgassing of the materials.
- **Compression** of the MLI during integration and contraction due to temperature changes.
- **Overlapping** of single MLI blankets, MLI seams and pressure areas due to deflection.
- Attachment and grounding methods.
- Number, size and location of **perforations**, which are used for better evacuation.

Generally the effectiveness of MLI decreases with the complexity of the configuration to be insulated, the number of MLI pieces, the reduced size of MLI blankets and the number of necessary overlaps. To illustrate this issue, Figure 4.3.18 compares theoretical values with different tested values of installed MLI.

In the development phase of a practical project the effectiveness of MLI is first assumed as having an **effective emissivity** between 0.01 and 0.03, with the higher value being selected for complex configurations. A wrong assumption of MLI effectiveness influences most notably the temperature level of the spacecraft. Based on results of the thermal vacuum test the temperature level may be adjusted by trimming the MLI in the area of radiators (increasing the radiator area lowers the temperature level, and vice versa).

The described MLI, which is optimized for vacuum conditions, cannot be used for missions to planets with their own atmospheres. For example, the atmosphere of Mars is composed of CO_2 gas at 6 to 10 mbar pressure. For this particular case, insulation with a **microporous molecular structure** of the type recently developed for terrestrial application may be applied (vacuum insulation panels for the building industry). The insulation is based on a powder-like, highly dispersed silica with particles having only point contact and additionally separated by a hardening agent (Figure 4.3.19). Consequently, solid body thermal conduction is very much reduced [4.3.13]. The dimensions of the remaining void volumes are smaller than the main free path of the enclosed low-pressure gas, which practically eliminates any gas-based heat conduction. In addition, radiative heat transport is minimized through application of powder-like infrared opacifiers.

In fact, **thermal tests** have confirmed that the heat conductivity remains at a low and constant value of barely 4 mW/mK up to a gas pressure of about 10 mbar (Figure 4.3.20).

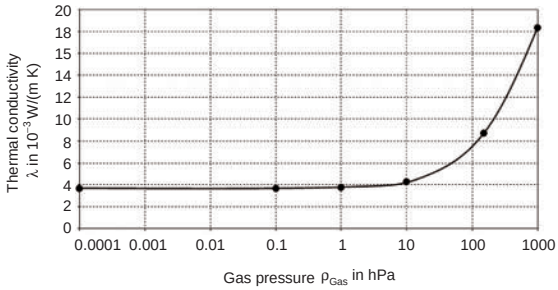


Figure 4.3.20: *Influence of gas pressure on the thermal conductivity of silica insulation [4.3.13].*

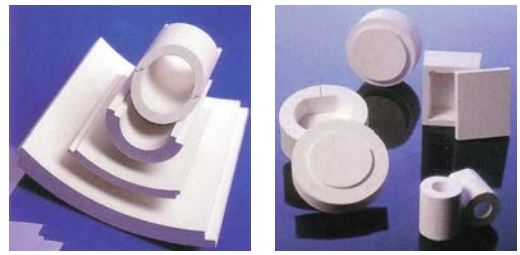


Figure 4.3.21: *Insulation shapes [4.3.13].*



Figure 4.3.22: *Insulation piece machined by milling processes.*

Compacting the material under pressure leads to plates and shapes which can be machined to arbitrary insulation pieces by milling and drilling (Figure 4.3.21). Figure 4.3.22 shows a base plate of an insulation system for a Mars probe [4.3.14].

4.3.4.3 Two-Phase Cooling Loops

In two-phase cooling loops the heat of vaporization of a liquid is transported between an evaporation and a condensation site. The simplest form of such a device is the **heat pipe**.

A heat pipe (see Figure 4.3.23) consists of a hermetically sealed, cylindrical tube with a capillary wick structure at the inner wall. After evacuating, the heat pipe is filled with a quantity of **fluid heat carrier** which just saturates the capillary structure. When heat is conducted into the evaporator and cooling is applied at the condenser, the generated vapor flows to the cooling site and condenses there by releasing heat. Because of the capillary forces of the capillary structure the condensed fluid is transported back to the evaporation site.

Heat pipes operate fully passively and without energy supply from outside. Because evaporation and condensation take place at about the same temperature, heat can be transported with very small temperature differences. In practice, temperature differences are observed, caused by radial flow resistance in the evaporation and condensation sites and due to pressure losses in the fluid flow. The **operating temperature range** of a heat pipe is theoretically limited

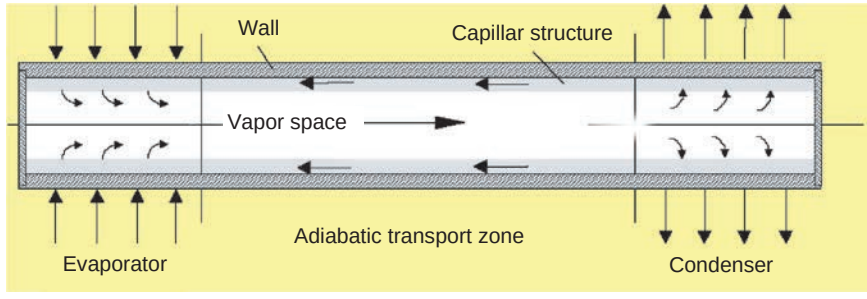


Figure 4.3.23: *Working principle of a heat pipe [4.3.15].*

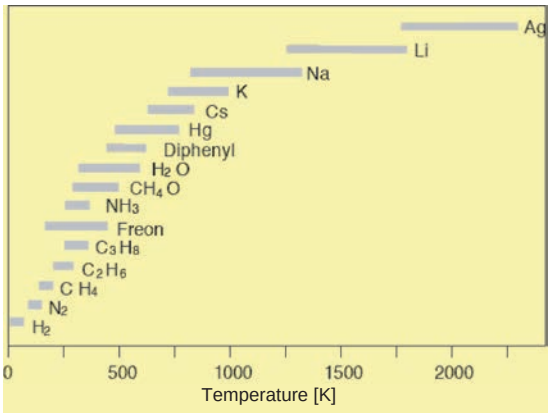


Figure 4.3.24: *Operating temperature ranges of some heat pipe liquids [4.3.15].*

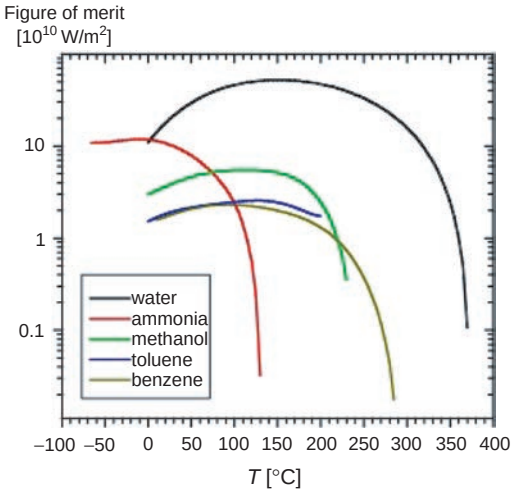


Figure 4.3.25: *Figure of merit for some heat pipe liquids.*

on the lower temperature side by the freezing point and on the higher temperature side by the critical point of the selected liquid. These ranges are given in Figure 4.3.24 for some liquids.

Further criteria for the effectiveness of a heat pipe are the thermohydraulic properties of the selected liquids. Liquids with high heat of vaporization κ, density ρ and surface tension τ, and with low dynamic viscosity η, are the most suitable. For assessment purposes, a **figure of merit** G has been introduced:

$$G = \frac{\kappa \rho \tau}{\eta} \qquad (4.3.36)$$

The figure of merit is plotted in Figure 4.3.25 for some liquids as a function of temperature. The diagram also shows the strong dependence of the liquid properties on the temperature. As a rule of thumb it can be stated that liquids in the medium-temperature range have their highest effectiveness at the normal boiling point (1 bar).

The selection of a **heat pipe liquid** is also governed by its **compatibility with the wall material** of the heat pipe tube. Material combinations which do not lead to corrosion and the creation of noncondensable gases are known from the literature. The compatibility data for materials in Table 4.3.5 is from [4.3.16].

The effectiveness of a heat pipe depends decisively on the **capillary pressure** which can be created with the selected capillary structure. Vapor and liquid establish a closed flow system in which the pressure gain through the capillarity is set against the pressure loss in the liquid and vapor flow. For all operational regimes the following is valid:

$$\Delta p_c = \Delta p_f + \Delta p_d \qquad (4.3.37)$$

Table 4.3.5: *Compatibility of heat pipe materials (Source: [4.3.16]).*

Wall material	Water	Acetone	Ammonia	Methanol
Copper	KE	KE	EE	KE
Aluminum	GEA	KL	KE	EE
Stainless steel	GEH	WK	KE	GEH
Nickel	WK	WK	KE	KL

KE = Compatibility verified by successful operation
KL = Compatibility confirmed in literature
WK = Probably compatible
EE = Use not recommended
KU = Compatibility not known
GEA = Gas generation at all temperatures
GEH = Gas generation at high temperatures.

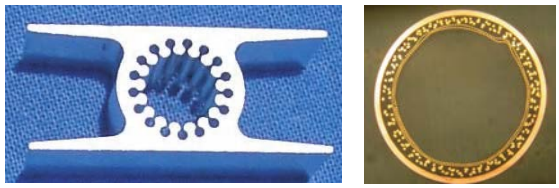


Figure 4.3.26: *Capillary structures: axial grooves (left), screen layers (right).*



Figure 4.3.27: *Heat pipe profile embedded into a sandwich structure.*

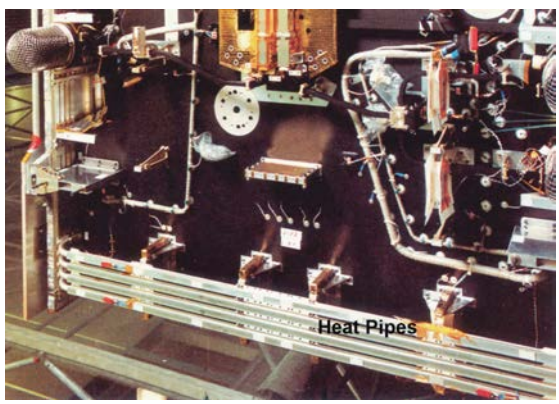


Figure 4.3.28: *Heat pipes connect the north and south sides of the German television satellite TV-Sat [4.3.17].*

where:
Δp_c = capillary pressure,
Δp_f = pressure loss in the liquid,
Δp_d = pressure loss in the vapor.

The capillary pressure Δp_c, that is the pressure difference at a curved liquid surface with a radius r, is generally defined by

$$\Delta p_c = 2\frac{\tau}{r} \qquad (4.3.38)$$

With an increase of heat transfer into the heat pipe, the velocities of the vapor and liquid flows increase and, accordingly, the pressure losses. The heat pipe fails when the flow discontinues; that is, when the pressure losses in the flowing fluid become larger than the pressure gain due to capillarity. Common **capillary structures** include open **axial grooves** or a combination of several **metallic screens** (Figure 4.3.26) which are attached to the inner wall of the heat pipe.

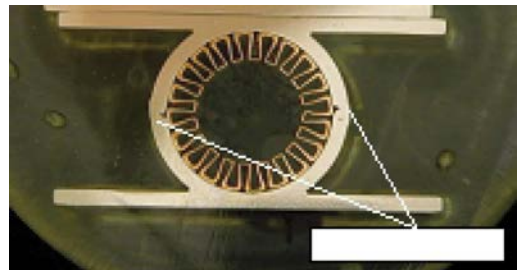


Figure 4.3.29: *Water–copper heat pipe.*

Axial grooves are advantageous since they can be machined with high reproducibility directly into the profile by an extrusion process. From known heat pipe materials, only some soft **aluminum** alloys are suitable for extrusion of axial grooved profiles. Aluminum, a commonly used lightweight structural material, is compatible with **ammonia** (Table 4.3.5). Heat pipes with this material combination operate in the temperature range of about –40 to +80 °C and combined with axial grooves they are the most common type of heat pipes in space projects. Applications are embedded heat pipes in radiator panels (Figure 4.3.27) and north–south connections in spacecraft (Figure 4.3.28).

In order to use thermal radiators more efficiently, payload components have been recently developed with operating temperatures of 100–200 °C. Since ammonia–aluminum heat pipes are not suited to this temperature range, the development of **water–copper heat pipes** has been initiated [4.3.18]. This material combination, with compatibilities as shown in Table 4.3.5, operates in the desired temperature range (see Figure 4.3.24). The proven axial grooves are manufactured using a galvanic process, where copper is electroplated on a negative form with external grooves. After dissolving the negative form in a chemical process, a thin-walled copper profile remains, which is inserted into an aluminum profile to increase the strength of the heat pipe (Figure 4.3.29). The two halves of

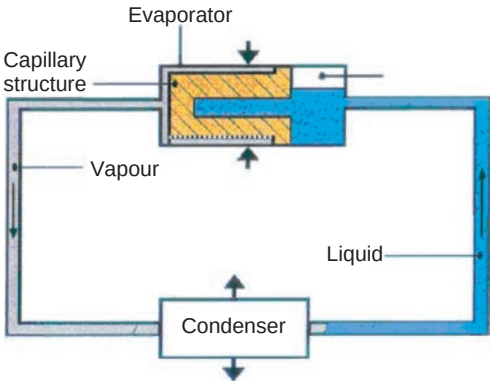


Figure 4.3.30: *Operating principle of a loop heat pipe.*

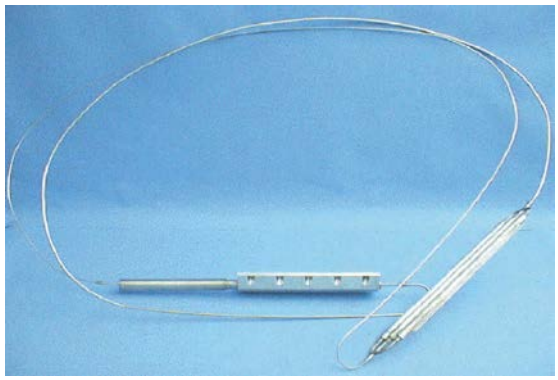


Figure 4.3.31: *Loop heat pipe (Source: TAIS Ltd).*

the aluminum profile are produced by an extrusion process and may be equipped with integrated flanges for interfacing to satellite structures.

The above-described heat pipes incorporate capillary structures along the full length of the profile, which must meet contradicting requirements. On the one hand, the dimensions which are effective for capillary action (pores, groove width) must be small to produce high capillary forces. On the other hand, these dense structures produce large pressure losses in the fluid flow. In practice an optimum has to be found during the heat pipe design phase. However, the resulting capillary force is in most cases very small and operation against the Earth's gravity is not possible. This feature leads to the need to accurately level the spacecraft during system tests.

The so-called **loop heat pipe** (LHP) avoids these constraints. According to Figure 4.3.30, the capillary structure in these units is located only in the evaporator. Under heat input in the evaporator the generated vapor flows in a smooth, small tube to the condenser, condenses there and returns as liquid in a second tube to a reservoir which is an integral part of the evaporator. Because the capillary structure is locally restricted to the evaporator, it is possible to apply very small pore sizes with high capillarity without producing undue high-pressure losses in the loop. Today's LHPs (Figure 4.3.31) are designed to operate with arbitrary orientation to gravity with heat transport capabilities of 1–2 kW.

Guiding the fluid flow through separated tubes also permits intelligent control of **vapor flow** as a function of evaporator temperature. As shown in Figure 4.3.32, a control valve can be integrated into the vapor line, which diverts the vapor flow into partial flows to the radiator and to the reservoir.

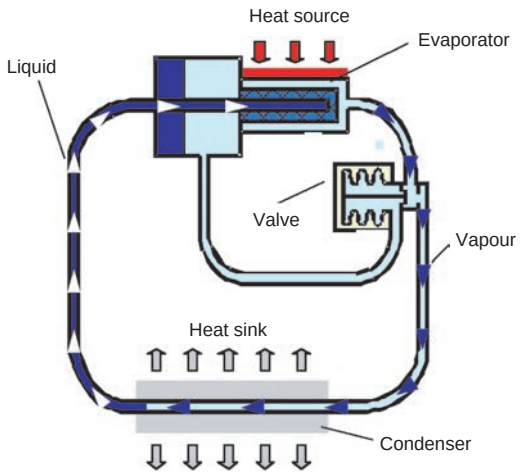


Figure 4.3.32: *Loop heat pipe with control valve.*

The **valve** (Figure 4.3.33) operates via an integrated spring bellows against a control pressure which can be adjusted to meet the control temperature selected at the evaporator. An increase of the evaporator temperature above a preselected control value causes a corresponding increase of the vapor saturation pressure and in turn a movement of the valve stem to open the valve outlet for an increased vapor supply to the radiator. As a consequence, the evaporator temperature decreases. In cases where the temperature falls below

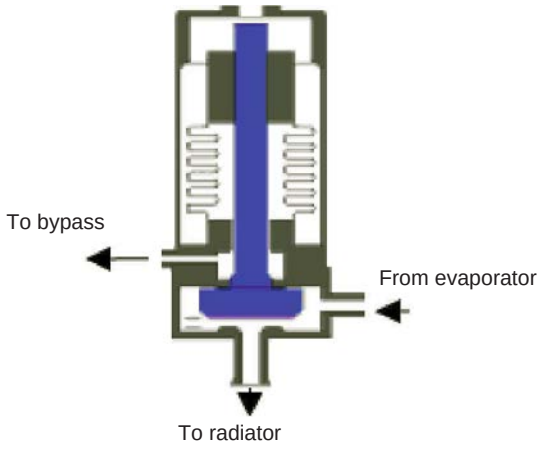


Figure 4.3.33: *LHP control valve.*

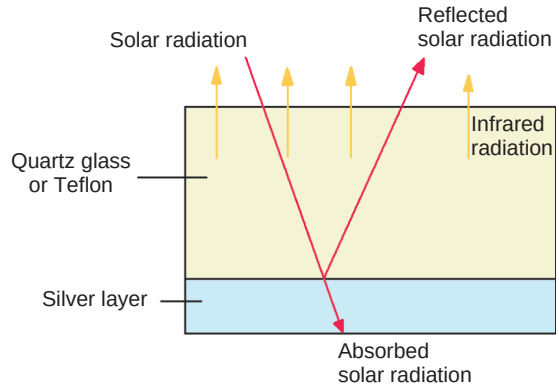


Figure 4.3.34: *SSM/OSR coating.*

4

the control value, a larger amount of vapor bypasses the radiator and flows back to the reservoir.

4.3.4. Thermal Surfaces

All visible surfaces of a satellite absorb or emit thermal energy and consequently influence the thermal control system. Also important are the **absorption of solar energy** and the **emissivity in the infrared band**. Inside the spacecraft all surfaces are usually coated with black paint ($\alpha = \varepsilon = 0.8$), in order to achieve a good heat exchange by radiation.

The external visible layer of the MLI consists of Kapton (DuPont's trade name for polyimide) with medium absorptance and emission properties. So-called **Beta cloth** (a beta-silica–fiber fabric) is applied in low Earth orbits, because Kapton would degrade due to atomic oxygen.

The thermo-optical properties of **radiator surfaces** have a predominant influence on the thermal control system. These surfaces should absorb low solar energy under direct solar radiation, but must exhibit at the same time a large emitting capability in the infrared range. So-called **second surface mirrors** (SSMs), which are also known as **optical solar reflectors** (OSRs), are employed. To achieve high emissivities they consist of translucent materials (quartz glass or Teflon) which are metallized on the backside with highly reflective silver or aluminum to decrease solar absorption (Figure 4.3.34).

Table 4.3.6: *α and ε values of commonly used thermal surfaces.*

Product	α		ε	
	Without aging	With (5 years)	Without aging	With (5 years)
Quartz SSM	0.08	0.2	0.8	0.8
Teflon SSM	0.08	0.13	0.78	0.75
White paint	0.26	0.44	0.88	0.88
Black paint	0.96	0.91	0.91	0.84
Polyimide (Kapton)	0.4	0.67	0.7	0.73
Beta cloth	0.24	0.35	0.88	0.88

Quartz SSMs are manufactured as small tiles, similar in size to solar cells (solar cell cover glass is also used for SSMs), whereas **Teflon SSMs** are installed as foil strips. Both types are attached to the radiator surface with acrylic or silicon adhesive. In addition to the information in Table 4.3.4, the α and ε values of the described surfaces with and without aging are summarized in Table 4.3.6.

To further improve radiator properties, **louvers** may be installed, the blades of which are operated with a **bimetallic mechanism** (Figure 4.3.35).

The bimetallic actuator contracts with increasing temperatures and the blades rotate into the open position. In this configuration the heat radiation increases. At reduced temperatures the blades rotate into the closed position and heat which radiates from the underlying surface is reflected by the polished blades.

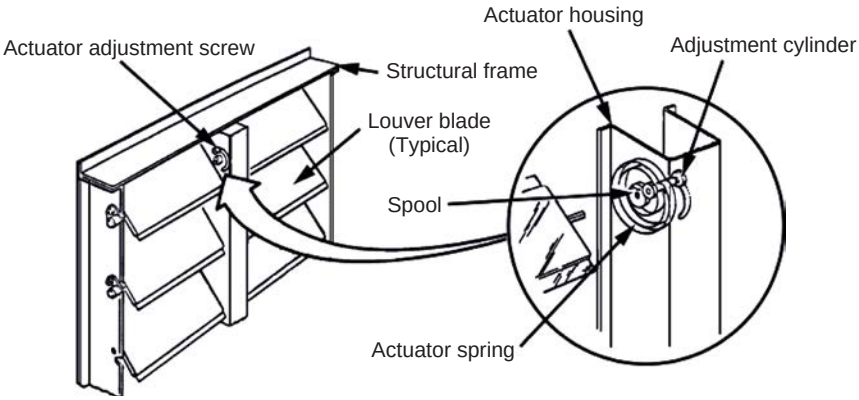


Figure 4.3.35: *Working principle of a thermal louver (Source: [4.3.19]).*

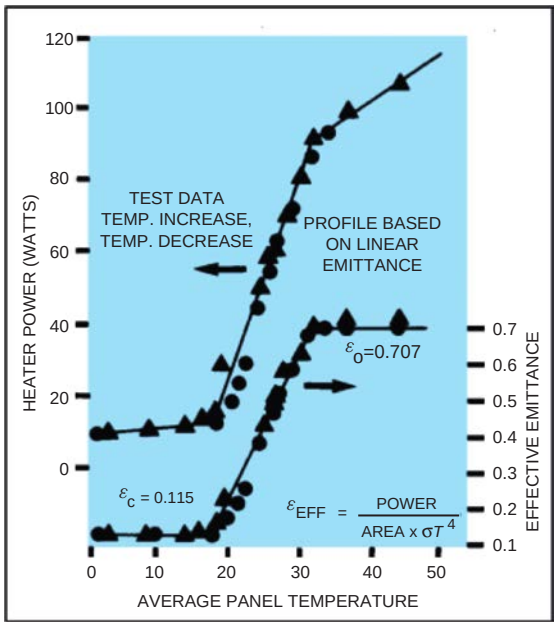


Figure 4.3.36: *Performance data of a thermal louver [4.3.19].*

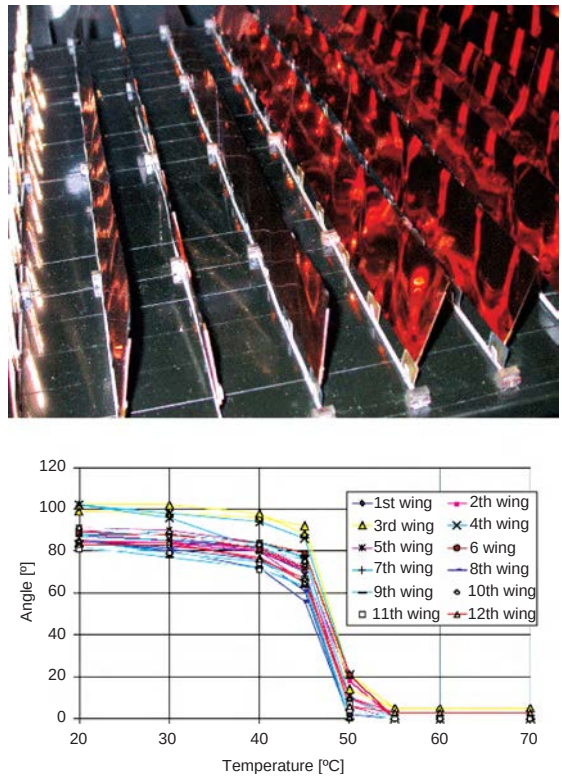


Figure 4.3.37: *Louver which closes at high temperatures [4.3.20].*

Louvers may be built for several temperature differences between the open and closed status. The differences are usually between 10 and 20 °C.

Typical performance values for a louver are given in Figure 4.3.36 [4.3.19]. The **effective emissivity**, shown in the figure, has been defined as the relationship between heat radiated from a surface with louvers and from an equivalent black surface. Accordingly, the maximum possible emissivity of an optimal surface decreases by 30% if the surface is equipped with louvers which are completely open.

Besides the louvers shown in Figure 4.3.35 and Figure 4.3.36, which open with increasing temperature, a new kind has been recently developed which opens at low temperatures to admit high radiative heat load from the radiator [4.3.20]. As shown in Figure 4.3.37, the angle of the louver blades with respect to the

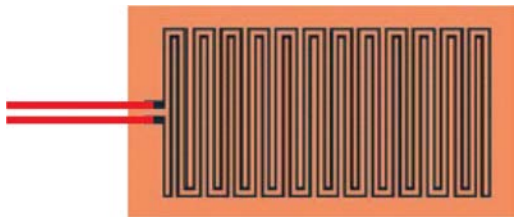


Figure 4.3.38: *Exemplary assembly of a foil heater.*

radiator surface is about 80°. At 45 °C the closing process starts, which is completed at about 55 °C. The specific emissivity according to the above-mentioned definition has been measured for this louver type to be 86.5%.

This design can be used for the east and west panels of a geostationary satellite, which are periodically exposed to direct solar radiation. In a practical application the east–west panels are equipped with dissipating equipment and connected via heat pipes. If one side is exposed to the Sun, the louver blades close to avoid solar absorption. The heat pipes transfer the equipment heat losses to the opposite, shadowed side, where the radiator surface with open louver blades radiates the heat into space. Since heat pipes operate in both directions, the thermal control concept is applicable to both spacecraft sides.

4.3.4.5 Heaters

Heaters are needed if a minimal thermal operating limit for a component must be absolutely maintained and if its dissipation heat is inadequate, or if the heat cannot be absorbed by the solar radiation. A heater is, in principle, an **electrical resistance**. Current is conducted through a thin wire with a specific resistance which converts electrical energy into heat. The heating wire, mostly consisting of copper, is embedded in thermoplastic foil for protection and insulation. Figure 4.3.38 shows an assembly of such a heating element.

According to their application, heaters are permanent in use, switched on/off by means of sensors, or electrically controlled by varying the current which maintains a constant component temperature. According to [4.3.21], heaters are used in the following subsystems:

- **Payload:** As a heat source to compensate for varying operation conditions.
- **AOCS (Attitude and Orbit Control System):** Sensors are generally mounted at thermally exposed places of the satellite; heaters are required to maintain the operating temperature.
- **TMTC:** Something similar (as for AOCS) applies to the antenna system of the Telemetry and Telecommand System.
- **Energy Supply:** Batteries must maintain very narrow temperature limits.
- **Propulsion System:** Hydrazine systems need temperatures > 9 °C, bipropellant and solid propellant engines at least 0 °C. The complexity of tanks, engines and conduits makes heaters necessary.

The heaters can be switched on and off by ground control. Temperature monitoring is performed via telemetry and the control of the heaters by telecommanding. However, they can also be controlled automatically on-board the satellite. The latter offers the following possibilities:

- Mechanical relays monitored by a microprocessor.
- Mechanical switches actuated by a bimetal sensor.
- Electric switches operated in a closed loop control with HCUs (Heater Control Units) or PCDUs (Power Control and Distribution Units).

An autonomous control loop normally has default settings for controlling the heater. Changing thermal conditions sometimes require adjustment of the settings by telecommanding.

4.3.5 Example of a Thermal Design

TerraSAR-X is a German remote sensing satellite whose primary payload is an active radar device. Its antenna acquires novel and high-quality X-band radar data of the entire Earth. The satellite orbits the Earth at an altitude of approximately 514 kilometers in a Sun-synchronous and nearly circular orbit. The choice of a dusk–dawn orbit (18:00 local time for the ascending node) provides good conditions for the design of the energy supply and thermal control systems: the solar array surfaces are directed almost vertically to the Sun, and eclipses occur only in the period around

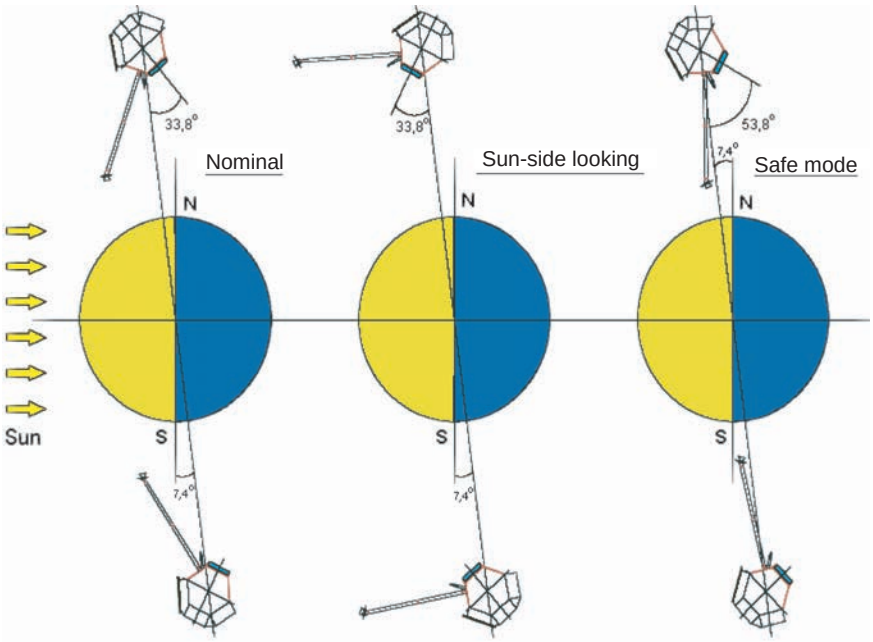


Figure 4.3.39: *The three main attitudes of TerraSAR-X within the orbital plane relative to the Sun: nominal, Sun-side looking and safe mode (Source: DLR).*

Table 4.3.7: *Definition of thermal load cases for the thermal design of TerraSAR-X (main thermal design cases).*

Parameter	Hot operation	Cold operation	Heater sizing
Purpose	Determination of maximum temperatures Radiator sizing	Determination of minimum temperatures	Heater sizing Determination of min. temperature
Point of time	03.01.	04.07.	04.07.
Sun declination	–23.4°	+23.4°	+23.4°
Solar constant	1428 W/m^2	1316 W/m^2	1316 W/m^2
Earth temperature	257 K	246 K	246 K
Albedo factor	0.35	0.20	0.20
Eclipse duration	–	22 min	22 min
Attitude	Nominal	Nominal	Safe mode
Unit dissipation	Hot timeline	Cold timeline	Safe mode

the summer solstice and have a maximum duration of 22 minutes.

The driving requirements for the thermal design of the satellite arise from the choice of a Sun-synchronous orbit and the different positions of the primary payload relative to the Sun (Figure 4.3.39):

- To maintain the temperature of the components and the radar antenna in the nominal attitude within the operational limits.
- To limit the heating of the radar antenna on the side exposed to the Sun (Sun-side looking) to a great extent.
- To minimize the heating power in the safe mode.

Variations in environmental conditions, orientations, operation durations or frequencies and the related dissipation rates of the bus and instrument units lead to the main thermal design cases summarized in Table 4.3.7. The hot or cold timeline defines the usage

frequency of the radar device. In the **hot timeline**, up to 70 data takes per revolution are to be made and for every data take a three-stage warm-up sequence must be run through. The hot operating design case determines the maximum temperatures and therefore the layout of the radiators in the nominal mode, whereas the very limited operation of electronic components in the safe mode serves for the layout of the heater system. According to technical properties, different **thermal operation limits** arise for bus and instrument units.

A distinction is made between operational, nonoperational and nominal limits. The **operational limits** mark the area in which electric or electronic assemblies can be switched on without being damaged. By the lower and upper **nominal operation limit** that temperature range is meant in which an electric or electronic instrument fulfills the functional requirements. The nominal limits usually correspond to the operational ones. The **nonoperational limits** apply to the disabled state and to mechanical components. For some electric or electronic assemblies, the latter are only relevant for storage and transport because they are switched on or active during the whole mission time, such as the on-board computer, the battery or the solar panel.

For the **heater layout** and especially for determining heater power within the scope of the thermal analysis, the control values are initially set 10 °C above the lower operational or nominal limit to cover uncertainties in the calculation. During operations, however, the heater control values are set typically to 3 °C above the minimum operations limit. Now the **thermal analysis** with the **node model** must clarify whether the temperatures are maintained in all nominal load cases and operational scenarios within the operational or nonoperational limits relevant for the thermal control. The bar diagram depicted in Figure 4.3.40 shows the result of this analysis for selected assemblies of the satellite that are in operation. The black bar covers the area of the thermal prediction, while the allowed temperature range is marked by the red bars. The green area shows the uncertainty span to be taken into account for temperature prediction. It can be seen that the uncertainty span of the lower limit is about 3 °C for heater-controlled assemblies because the temperature can be actively maintained. An uncertainty span of 5 °C applies for the remaining components if their thermal behavior was sufficiently verified by testing. Otherwise, 10 °C has to be allowed. An overheating scenario is excluded, although too long or too frequent use of the radar antenna might cause overheating.

Together with the thermal operational limits, some components of course also have temperature stability and spatial gradient requirements.

Figure 4.3.40 and Figure 4.3.41 show exemplarily the **predicted temperature variation** of single thermal nodes in the hot load case over three revolutions. In Figure 4.3.40, the nodes represent electronic components of the radar device and in Figure 4.3.41 battery modules and parts of the PCDU.

One can clearly recognize the strong temperature variation of those units which are switched on and off in the time interval. However, components which are switched on permanently are characterized by a nearly constant temperature variation. In addition, the temperature correlation is clearly visible between switchable units and the battery or the PCDU. The enabling of electronic components leads to increased activity of the PCDU and to a charge and discharge of the battery which results in the generation of additional heat in both cases.

On the basis of the load cases, the given structure, bus and payload instruments and the thermal analysis, the following detailed thermal configuration was designed for TerraSAR-X:

- **Thermal insulation** of the satellite by MLI and suitable surface coatings at bus and instrument levels.
- **Heat radiation** into space by the use of **foil radiators** as shown in Figure 4.3.43. The surface ratio of MLI to foil radiators is chosen so that the average temperature of the electronic units does not exceed 30 °C in the hot load case.

The external layer of the MLI exposed to space consists of a Kapton foil coated with SiO_2 (yellow area) with a ratio of $\alpha/\varepsilon = 0.34/0.62$. MLI in the visibility area of the star sensors (black surface) has a black coating ($\alpha/\varepsilon = 0.95/0.90$) to minimize stray light effects. The radiators (red surface) consist of an unperforated Teflon foil silvered on the inner side. It has a thickness of 127 µm and a ratio of $\alpha/\varepsilon = 0.12/0.75$. Thermally conducting silicone–elastomer insulation is inserted

Item	Lower limit	Predicted low	Predicted high	Upper limit
TRM	-20	-6,0	35,9	60
PPC	-25	-4,8	27,9	50
PCFN	-20	-3,9	24,6	60
WG	-80	-7,5	29,9	80
LAA	0	24,4	27,8	50
DCE	-10	4,4	41,5	50
RFE	-10	12,1	35,9	50
ACE	-10	0,3	30,1	50
SSMM	-20	6,1	24,1	40
IPDU	-25	-1,2	35,3	45
RPS	-25	0,8	32,6	60
XDL antenna	-135	6,3	33,0	150
MOD	-15	8,6	29,6	55
TWT	-15	6,5	41,0	70
EPC	-15	8,7	30,2	55
WGS, ISO & FILT	-15	2,3	33,7	55
LCT HP I/F	-25	11,3	19,3	25
IGOR Receiver	-10	16,8	36,5	50
IGOR LNA	-30	6,1	26,5	50
POD antennas	-90	-40,4	-0,3	100
ROA antennas	-90	-14,2	12,0	100
LRR	-50	-9,3	19,1	50
Structure	-100	-6,6	35,7	110
XDL antenna boom	-135	3,0	19,0	100
OBC	-25	-11,7	33,3	50
RFEA	-20	-7,3	35,3	50
S-Band antennas	-100	-36,6	69,9	100
RW	-10	1,2	33,2	50
MTQ	-40	2,6	27,3	60
MAG	-30	3,8	25,6	50
STE	-20	8,4	33,1	50
STH	-40	-31,7	-26,7	20
GPS antennas	-110	-27,7	35,0	110
GPS filter / amplifier	-30	-17,6	26,5	50
IMU	-40	2,0	5,8	60
PROP	10	13,8	34,0	45
Solar array	-80	-51,0	128,6	130
PCDU	-25	-2,3	37,4	55
Battery	0	5,5	26,5	35

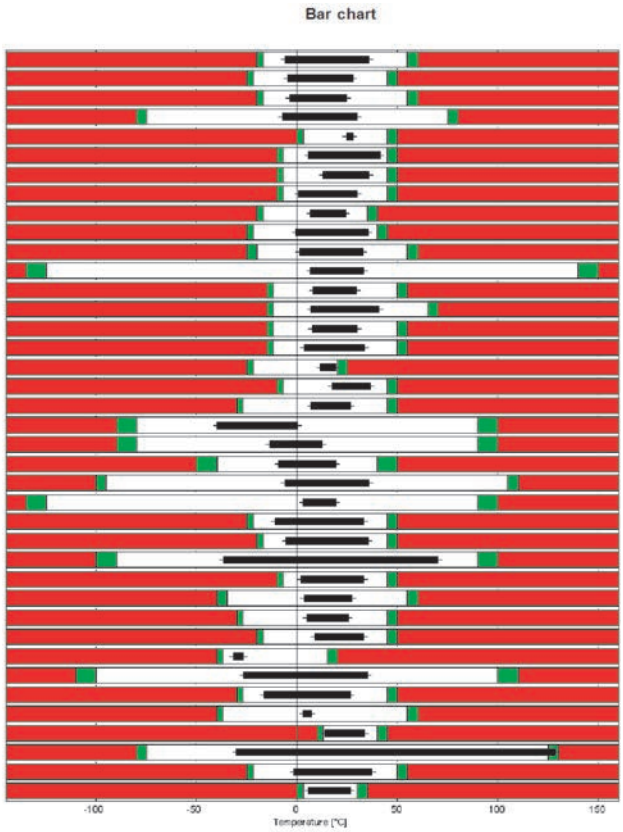


Figure 4.3.40: *Operational limits of TerraSAR-X assemblies with the predicted temperature ranges (black bars).*

between highly dissipating electronic units and their bearing structure to improve contact heat transfer.

- Use of **heat pipes** to achieve high heat transport.

The secondary payload of the satellite, the laser communication terminal (LCT), generates so much heat during operation that it must be transported via heat pipes to a dedicated radiator. The LCT radiator is thermally decoupled from the satellite structure by using an MLI layer inside the radiator and insulation elements at the assembly points.

- Implementation of an **electric heater system** for temperature control and stabilization of the (payload) instruments and other components. The heater system consists of independently controllable heater lines, as depicted in Figure 4.3.44.
- **Control** of the heater lines by means of software and corresponding temperature sensors (thermistors).

Temperature regulation is controlled by the OBC (On-Board Computer). The measured temperatures are compared to the commanded threshold settings of the active heater lines, and the heaters are either switched on or off by the PCDU. For safety reasons, nearly all heaters and their switching circuits are so distributed that if a single heater or a whole heater group fails, another single heater or group can take over the heating function, which also represents **redundancy**. Nearly every heater in module 1 thus has a partner in module 2. Every heater group is protected and controlled by a **latch current limiter** (LCL), and a

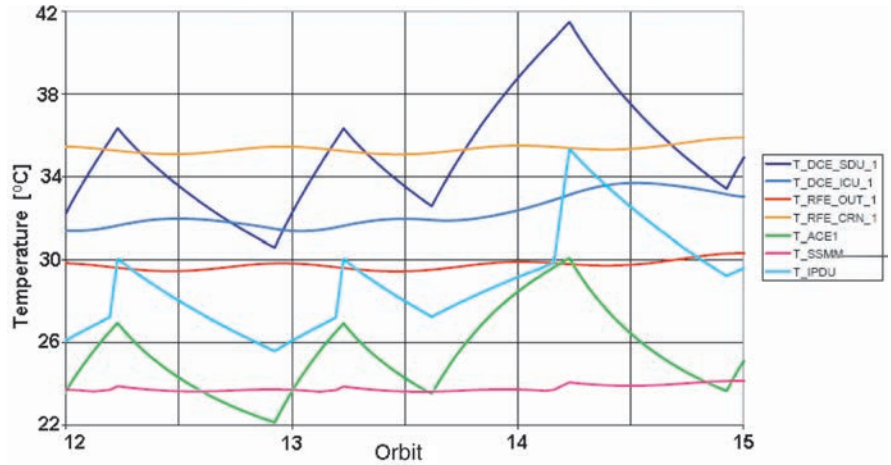


Figure 4.3.41: *Temperature profiles of thermal nodes representing the radar equipment in the hot load case (Source: DLR).*

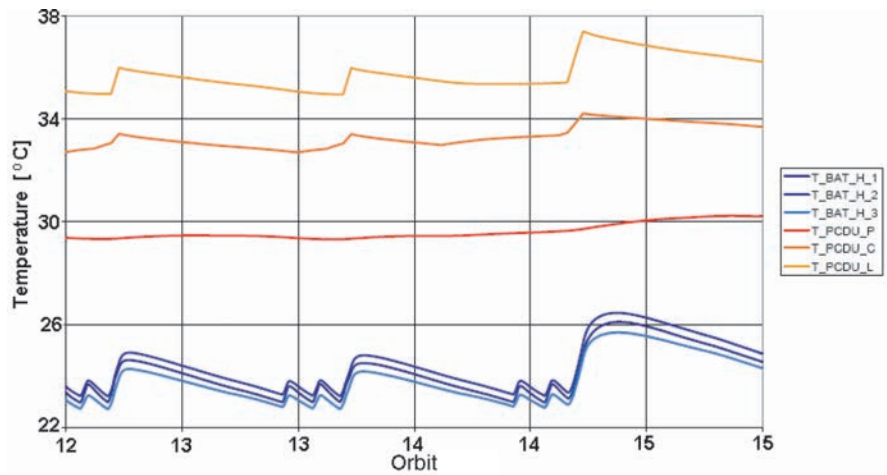


Figure 4.3.42: *Temperature profiles of thermal nodes representing the battery and PCDU in the hot load case (Source: DLR).*

nominal as well as a redundant thermistor is assigned to each heater line for redundancy in case of a failure. In all, 150 sensors are used: 114 for temperature monitoring and control, and 36 for the attitude control system that are part of the CESS (Coarse Earth and Sun Sensor). Along with these 150 thermal sensors at bus and instrument levels, there are special instrument thermal sensors used for calibration purposes. Because of this function they are assigned rather to the instrument bus and calibration system.

Although the heater thermal control system is operated autonomously by the OBC, routine monitoring of the temperatures is possible. Operational and technical constraints demand from time to time active intervention in the controlling process from the ground. Therefore, a suitable monitoring system and configuration software are required, which will be explained in more detail in the next section.

4.3.6 Operation of the Thermal Control System

The components of the propulsion and attitude control system, the energy supply system and in particular the electronic components of the radar device need an operating temperature for their functionality and performance that is possibly still not reached after

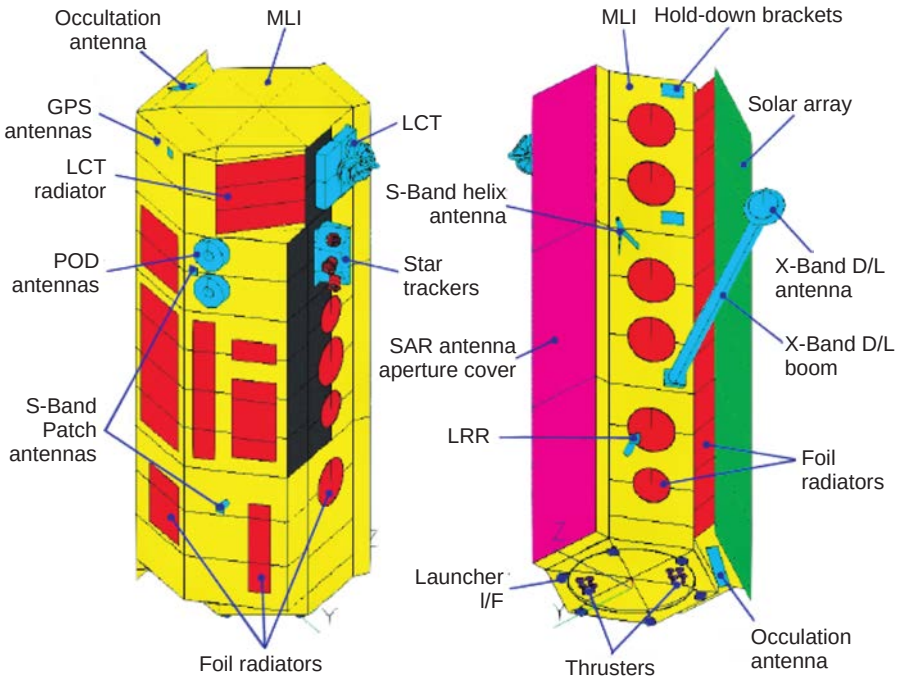


Figure 4.3.43: *Outer configuration of MLIs and foil radiators for TerraSAR-X (Source: DLR).*

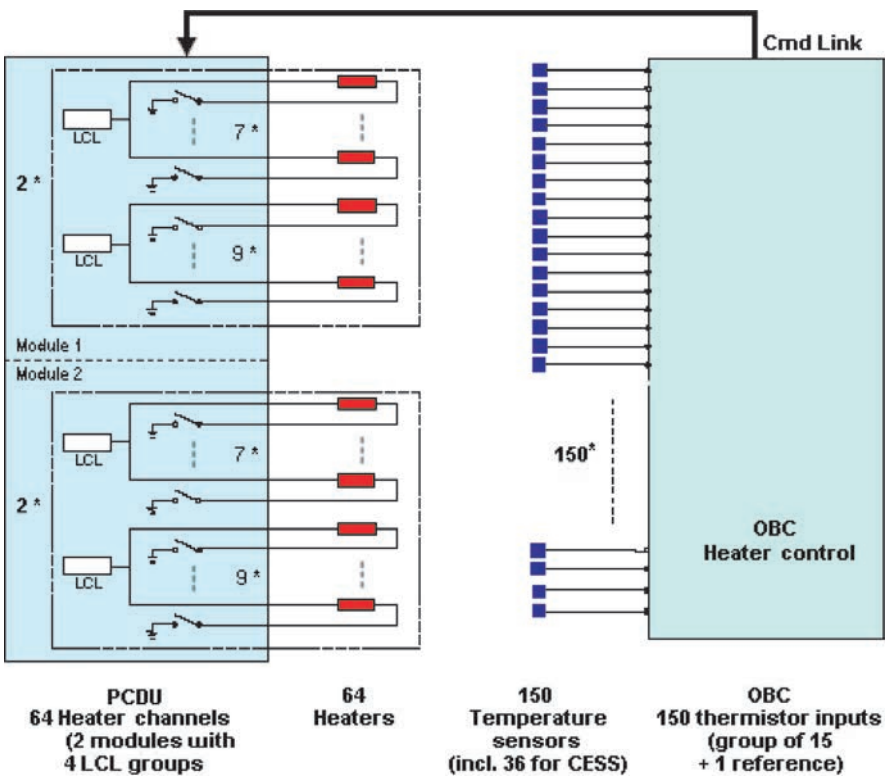


Figure 4.3.44: *Functional layout of the active heater control system of TerraSAR-X (Source: DLR).*

separation from the launch vehicle. Therefore, it is mandatory that selective and stepwise **warming** of the subsystem and payload components be performed in the LEOP (Launch and Early Orbit Phase). Theoretically, all available heaters could be switched on immediately and remain switched on until all temperatures are in the nominal operating range. However, this would overload the battery, especially in the cold state, and cause a very quick discharge which would again endanger the boot sequence of the subsystem's essential assemblies (OBDH, On-Board Data Handling; AOCS, Attitude and Orbit Control System). In addition, the dissipating heat of the instruments can be utilized to conserve heat power or energy. For this reason, the following **warm-up steps** with the corresponding configuration tables have been defined for TerraSAR-X for the heater settings:

- Table 1 with Survival/Safe limits (SURV)
- Table 2 with Standby limits (NOM1)
- Table 3 for Radar/XDA operation with operational limits (NOM2)
- Table 4 for Radar/XDA operation with nominal limits (NOM3).

In the **SURV table**, the threshold settings of the propulsion system are increased to maintain the temperature of the flow valves for thrusters and propellant sufficiently above the lower limit. The threshold settings in the **NOM1 table** initiate a warming up of the battery and AOCS assemblies. The settings for the thruster valves and propellant are, however, decreased again. Only in the **NOM2 table** are the operational threshold settings for the payload-relevant components raised. In the **NOM3 table**, nominal threshold settings are defined to ensure better performance of the radar instrument and the X-band Downlink Assembly (XDA).

The use of **configuration tables** for the control of all 64 heater lines considerably simplifies the operations on the ground because only one single command must be executed to activate the table. The table can be selected by choosing a formal parameter. Every table contains the definition of the control thermistor, the control status, on and off thresholds, as well as the control frequency (LF, Loop Frequency) for every heater line or for every row (line). The control frequency can be set to either 1/32 or 1/16 Hz by commanding. The lower frequency is assigned to group 0 and the higher one to group 1.

The monitoring of the (sub)systems is done by means of a special **telemetry system** which processes the measurements or housekeeping data and shows them on alphanumeric, graphic and synoptic display pages (combined display of graphic objects, texts and numeric values) (see Section 6.2). There are two alphanumeric pages for the display of the heater line settings. Figure 4.3.45 shows the first page on which the settings of the first 32 heater lines of the configuration table NOM1 are displayed in tabular form.

To clearly arrange the heater line activities for the flight control, a display has been designed by the thermal control engineer showing the switch status of the heaters and the measured temperatures synoptically and groupwise. Figure 4.3.46 shows the instantaneous status of the heater control system in the thermal control mode NOM1. When the PCDU switches on several heaters the corresponding graphic object becomes blue.

For clarity, bus and instrument thermistors are split over two displays and again grouped into subsystems, assemblies or instruments. Figure 4.3.47 shows the display page for the temperatures of bus thermistors.

It is often necessary to monitor **temperature variations** for a short period on a real-time basis. Therefore, graphic display pages have been designed which plot the temperature variation of selected thermistors as a function of time. Figure 4.3.48 shows the typical temperature variation at the battery which is caused by the control process, that is by switching the heater on and off.

The recorded run of the control process in Figure 4.3.48 is simulated. In reality, only 5 to 10 minutes are available during contact for real-time telemetry.

Sometimes there are **contingency situations** requiring the thermal control engineer to react appropriately to recover from a failure; that is, to find solutions or workarounds so that the thermal control system operates properly again. This relates to the electric heater system because interventions are only possible from the ground. The delamination of a part of the MLI from the structure would be a possible scenario which could cause a thermal problem. However, direct intervention is not possible. **Contingencies** in the heater system can

TCS5401A - Table Settings of Heater Lines 01-32

Dumped TC table NOM1

Line	Thermistor ID	Control status	ON threshold	OFF threshold	LF ID
01	C_11_FCV-1A_N	ENA	16.0 °C	16.5 °C	1
02	C__7_STH-1_N	ENA	-35.1 °C	-34.5 °C	0
03	C__4_DCE-1	ENA	-35.1 °C	-34.5 °C	0
04	C_38_RW-3	ENA	-5.0 °C	-4.5 °C	0
05	C_49_XFE-P4	ENA	-35.1 °C	-34.5 °C	0
06	C__8_BAT-1	ENA	15.0 °C	15.4 °C	0
07	C__0_REF_G1	DIS	-160.0 °C	-160.0 °C	0
08	C_10_LCTFUS_N	ENA	-30.0 °C	-29.6 °C	0
09	C_12_FCV-2A_N	ENA	16.0 °C	16.5 °C	1
10	C_55_STH-1_R	DIS	-160.0 °C	-160.0 °C	0
11	C_57_TANK_N	ENA	16.0 °C	16.5 °C	0
12	C132_SSMM	ENA	-35.1 °C	-34.5 °C	0
13	C134_STE-1	ENA	-15.0 °C	-14.5 °C	0
14	C_18_XFE-P8	ENA	-35.1 °C	-34.5 °C	0
15	C117_IGOR	ENA	-5.0 °C	-4.5 °C	0
16	C_52_RFE-1	ENA	-35.1 °C	-34.5 °C	0
17	C_43_FCV-3A_N	ENA	16.0 °C	16.5 °C	1
18	C148_OBC	ENA	-20.0 °C	-19.5 °C	0
19	C__3_LAA-1_N	ENA	-35.1 °C	-34.5 °C	0
20	C130_XFE-P12	ENA	-35.1 °C	-34.5 °C	0
21	C__9_PCDU	ENA	-20.0 °C	-19.5 °C	0
22	C_41_PROP_FL-A	ENA	16.0 °C	16.5 °C	0
23	C__0_REF_G1	DIS	-160.0 °C	-160.0 °C	0
24	C__5_MOD_P	ENA	-20.0 °C	-19.5 °C	0
25	C_44_FCV-4A_N	ENA	16.0 °C	16.5 °C	1
26	C_23_STH-3_N	ENA	-35.1 °C	-34.5 °C	0
27	C135_RFEA	ENA	-15.0 °C	-14.5 °C	0
28	C_35_LAA-2_N	ENA	-35.1 °C	-34.5 °C	0
29	C147_IPDU	ENA	-25.0 °C	-24.6 °C	0
30	C145_ACE	ENA	-35.1 °C	-34.5 °C	0
31	C158_XFE_PPC12	ENA	-45.0 °C	-44.5 °C	0
32	C_42_LCTHP_N	ENA	-20.0 °C	-19.5 °C	0

Figure 4.3.45: *Alphanumeric display page showing the settings of the first 32 heater lines (Source: DLR).*

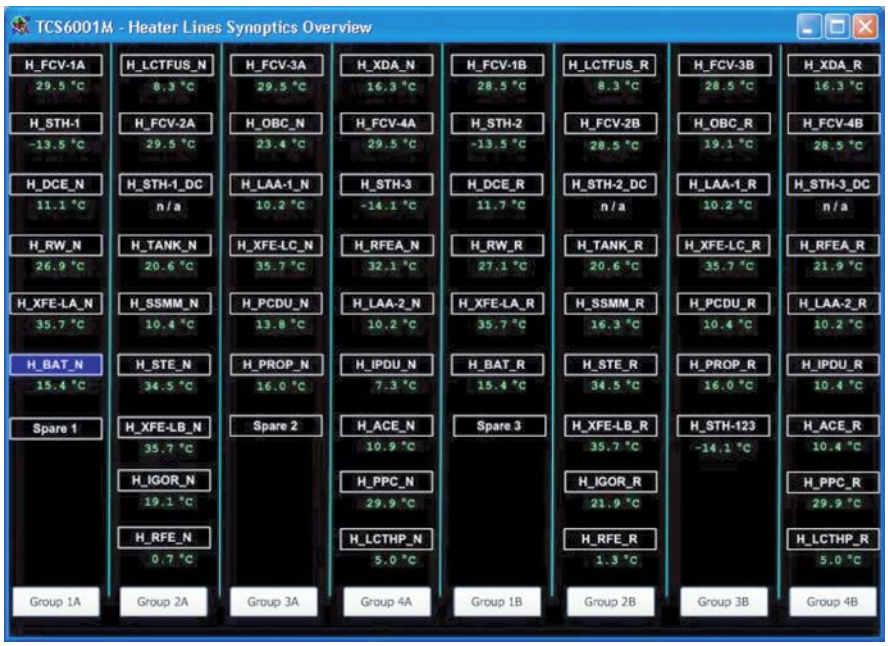


Figure 4.3.46: *Synoptic display page which clearly organizes all heater line activities (Source: DLR).*

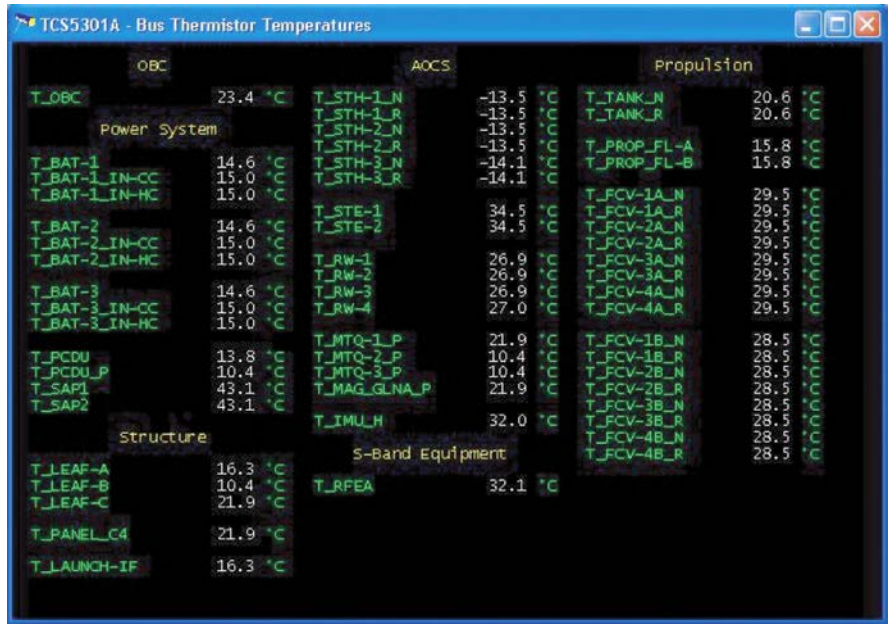


Figure 4.3.47: *Alphanumeric display page showing the measured temperatures at the bus level (Source: DLR).*

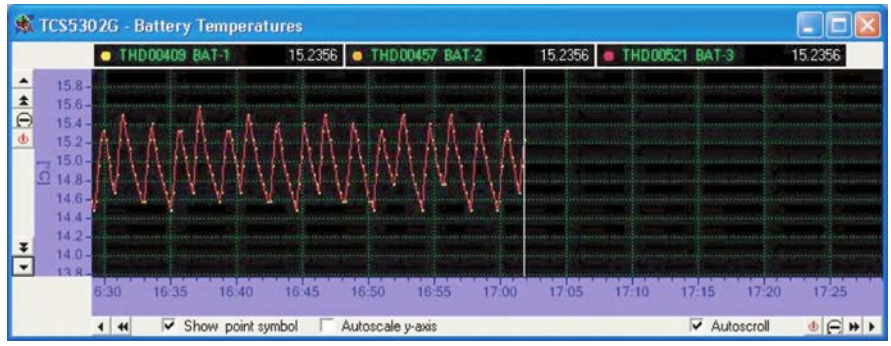


Figure 4.3.48: *Graphic display page showing battery temperature variations as a function of time (Source: DLR).*

arise if thermistors and/or heaters fail. **Faulty thermistors** can be identified because they show the wrong temperature values. The reasons normally are:

- Delamination of the sensor from the surface
- Cable breakage
- Short circuit.

A cable break is indicated by the maximum value of the temperature calibration curve and a short circuit by the minimum value.

A **failed heater** can be recognized if it is either permanently switched on or cannot be switched off. The reasons are:

- Delamination of the heater from the surface
- Open switching circuit due to a defective heater
- Constantly closed switching circuit due to malfunction of the PCDU
- Short circuit.

If a heater delaminates from the surface, the threshold is mostly not reached because the heat can no longer flow into the assembly to be heated. If the heater or the switching circuit is defective then there is no current. When it falls below the threshold, the displayed current or heater power remains zero, although the OBC has commanded "ON." If a heater cannot be switched off by commanding, the malfunction is attributed to a faulty PCDU.

Bibliography

[4.3.1] ECSS-E-30 Part 1B. *Space Engineering, Mechanical – Part 1: Thermal Control*, 2007.

[4.3.2] ECSS-E-10-04A. *Space Environment*, 2000.

[4.3.3] Ley, W., Hallmann, W. *Handbuch der Raumfahrttechnik*, 2. Auflage. Munich: Carl Hanser Verlag, 1998.

[4.3.4] Anderson, B.J., Smith, R.E. Natural Orbital Environment Guidelines for Use in Aerospace Vehicle Development. *NASA TM 4527*, June 1994.

[4.3.5] Spacecraft Thermal Control. *NASA SP-8105*, May 1973.

[4.3.6] Gilmore, D.G. (ed.) *Spacecraft Thermal Control Handbook, Volume 1: Fundamental Technologies*, Second Edition. Reston, VA: AIAA, 2002.

[4.3.7] Welch, J.W., Ruttner, L.E. An Experimental and Computational Analysis of the Thermal Interface Material Calgraph. 24th Thermophysics Conference, Paper AIAA-89-1658, 1989.

[4.3.8] Hallmann, W. Die Bedeutung der Formfaktoren in der Raumfahrt-Thermodynamik und ihre Ermittlung mittels Formfaktometer. *Z. Flugwiss. Weltraumforsch.*, **2**, Heft 2, 1978.

[4.3.9] ECSS-E-30 Part 1A. *Space Engineering, Mechanical – Part 1: Thermal Control,* April 2000.

[4.3.10] ESATAN Engineering Manual, *EM-ESATAN-056*, October 2006.

[4.3.11] Hallmann, W. Thermische Entkopplung eines Schrägkugellagers im Hochvakuum. *Vakuum-Tech.*, **24** Jahrgang, Heft 3, 1975.

[4.3.12] Angaben des Herstellers, Austrian Aerospace.

[4.3.13] Prospektmaterial des Herstellers Wacker Chemie GmbH. für das Produkt Wacker WDS®

[4.3.14] Romberg, O. *et al.* Netlander Thermal Control. *Acta Astronaut.*, **59,** 946–955, 2006.

[4.3.15] Stephan, P. Höhere Wärmeübertragung – Verdampfung und Kondensation. Vorlesung, Technische Universität Darmstadt.

[4.3.16] Dunn, P.D., Reay, D.A. *Heat Pipes,* Fourth Edition. Oxford: Pergamon Press, 1994.

[4.3.17] Fischer, H.-M. *Europäische Nachrichten-Satelliten, Von Intelsat bis TV-Sat.* Lemwerder: Stedinger Verlag, 2006.

[4.3.18] Schlitt, R. *et al.*: Development of a Light-Weight Copper/Water Heat Pipe. 14th International Heat Pipe Conference, Florianópolis, Brazil, April 22–27, 2007.

[4.3.19] Thermal Control Louvers. Orbital Technical Service Division, undated.

[4.3.20] Reichenberger, K., Matovioc, J. Development, Manufacturing and Verification of Micro-Electro-Mechanical Louvers. undated.

[4.3.21] Renner, U., Nauck, J., Balteas, N. *Satellitentechnik: Eine Einführung.* Berlin: Springer Verlag, 1988.

4.4 Satellite Propulsion

Hans Dieter Schmitz

The information in this section about propulsion systems for Earth-orbiting satellites also applies to **interplanetary probes**, since there are only slight differences. Similar systems are also used on launcher upper stages for roll and attitude control. Significant differences will be mentioned in the following sections.

4.4.1 Fundamentals of Satellite Propulsion

The first satellites, such as Sputnik or Explorer, were not equipped with propulsion systems. They were stabilized by rotation around their main axis, which was sufficient for the tasks they performed. But for the first geostationary communications satellites, propulsion systems were required for orbit and attitude control.

4.4.1.1 Propulsion System Tasks

Depending on the type and mission, propulsion subsystems have the following tasks [4.4.1], [4.4.2]:

- **Apogee Injection:** The launcher carries the satellite into an elliptical transfer orbit. In order to reach the final circular geostationary orbit, the apogee engine with a thrust level of 400 to 600 N is activated in the apogee to move the satellite with three or four orbit correction maneuvers into the required circular orbit.
- **Orbit Control:**
  - Drift initiation in the orbit plane in order to reach the target position (longitude) above Earth
  - Maintenance of the required position, east–west station keeping
  - Inclination control, that is the removal of an inclination of the orbit relative to the equatorial plane, north–south station keeping
  - Injection into a "graveyard orbit."

  Orbit control is achieved by means of thrusters of 10 to 22 N.

- **Attitude Control:**
  - Orientation of the satellite for pointing antennas toward the Earth or the solar arrays toward the Sun
  - Active stabilization around all three satellite axes as well as removal/control of attitude disturbances during or after orbit control maneuvers
  - Maintenance of the spin rate of spin-stabilized satellites as well as the control of the spin rates of reaction wheels.

  Depending on the size of the satellites or probes, attitude control is performed with thrusters having thrust levels between 1 and 22 N.
- The propulsion systems of interplanetary probes have to perform additional tasks such as:
  - Precise **course corrections** during flights lasting several years
  - **Braking maneuvers** (deceleration) to swing into an orbit around a moon or planet.

Many different **disturbances** require attitude and orbit control for satellites and probes, such as disturbances of satellites in low Earth orbits caused by the residual atmosphere, disturbances caused by the magnetic field of the Earth or Moon, disturbances generated by the movement of solar arrays, misalignment of orbit control thrusters, sloshing of propellants in the tanks, etc.

4.4.1.2 Stabilization Methods

The method used to stabilize a spacecraft has a major influence and impact on the design and configuration of the propulsion system. A distinction is made [4.4.4] between spin stabilization and three-axis active stabilization.

A spinning satellite rotates around its main axis at about 90 to 120 rpm, but lower spin rates of a few revolutions per minute have been used as well. Communications satellites such as Intelsat III to VI or the European Meteosat weather satellites were and are spin stabilized.

The thrusters for orbit and attitude control are oriented either **radially** (acting through the center of gravity of the satellite) to be operated in pulsed mode, or **axially** (with an offset to the spin axis) to be operated in steady-state mode, if the spin rate is sufficiently high.

A **three-axis actively stabilized satellite** – today's preferred stabilization method – requires special positioning and control effort for the orbit and attitude thrusters, but on the other hand allows the placement of antennas, instruments or solar arrays outside the central spacecraft body. Stabilization is achieved with reaction wheels in combination with the propulsion system, which controls the spin rate of the reaction wheels. Attitude control is described in more detail in Section 4.5.

4.4.2 Propulsion System Types

The most frequently used propulsion systems are cold gas, monopropellant, bipropellant and, most recently, electric propulsion systems. The selection of the propulsion type depends to a great extent on the type, size and mission of the spacecraft. Characteristic parameters are mass, propellant demand or "delta-v," reliability and mission lifetime. All figures shown in Sections 4.4.2.1 to 4.4.2.3 below are simplified sketches [4.4.1], [4.4.2], [4.4.4].

4.4.2.1 Cold Gas Propulsion

In the 1960s cold gas propulsion systems were preferred as standard reaction control systems [4.4.1], [4.4.2]. Cold gas propulsion is clearly the **simplest** and today the **most cost-effective** type of propulsion. Inert, nontoxic gases stored under high pressure are used. Figure 4.4.1 (left) presents a simplified flow schematic consisting of a gas tank, pressure regulator and cold gas thrusters.

In general, nitrogen or argon is used because of their low molecular weight. The pressure is reduced to the operating pressure of the cold gas thrusters by the regulator. The thruster itself consists of a solenoid valve and a nozzle, using the gas as propellant in pulse or steady-state mode. The most important task of cold gas propulsion systems on-board satellites is attitude control. For very small or minisatellites compact propulsion systems using butane as propellant have been developed and flown.

For attitude control of launcher upper stages the gas in the launcher propellant tanks (e.g., a helium/

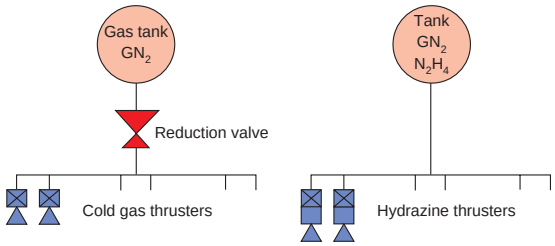


Figure 4.4.1: *Schematic of a cold gas (left) and a monopropellant (right) hydrazine propulsion system.*

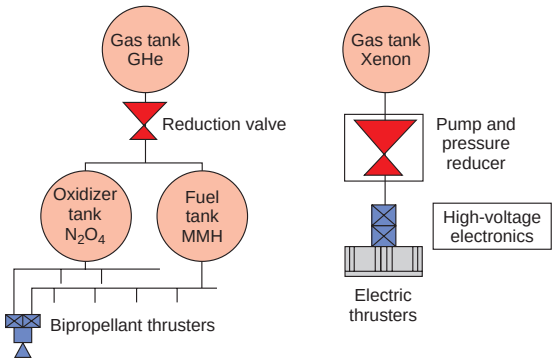


Figure 4.4.2: *Schematic of a bipropellant system (left) and an electric propulsion system (right).*

hydrogen mixture) is used as propellant for attitude control purposes as well.

4.4.2.2 Monopropellant Propulsion

With increasing satellite masses, propulsion systems of higher performance were required to reduce the propellant mass. Monopropellant propulsion systems – initially **hydrogen peroxide** (H_2O_2) and, after a suitable catalyst became available, **hydrazine** (N_2H_4) – led to the desired performance increase by a factor of 2 to 3 compared to cold gas [4.4.1], [4.4.2]. Figure 4.4.1 (right) presents the simplified flow schematic of a hydrazine propulsion system.

In the simplest case, propellant hydrazine and pressurant nitrogen or helium are stored in the same tank, separated by a rubber diaphragm. The tank is connected to one or more thrusters using an appropriate tubing system, the thruster consisting mainly of a flow control valve, a heat barrier with injector, a catalyst bed and an expansion nozzle. Due to the higher performance (compared to cold gas), monopropellant propulsion systems are used for **orbit and attitude control**; hydrazine thrusters, for example, can be operated in steady-state mode as well as in pulse mode to deliver the required low impulse bits.

4.4.2.3 Bipropellant Propulsion

The next step in the evolution of satellite propulsion entailed higher costs and increased complexity, but compared to monopropellant systems yielded 25 to 30% higher performance. Initially only considered for north–south station keeping and **apogee injection**, bipropellant propulsion systems are now also used for **attitude control** as well because of the successful development of thrusters with low thrusts between 10 and 22 N [4.4.1], [4.4.2], [4.4.4].

Bipropellant systems according to Figure 4.4.2 (left) need at least two propellant tanks, one each for **oxidizer** and **fuel**, as well as at least one **high-pressure helium tank** with a pressure regulator. Bipropellant thrusters consist of two flow control valves (to control oxidizer and fuel flow), an injector, a combustion chamber and a nozzle.

4.4.2.4 Electric Propulsion

Thrusters for electric propulsion systems use **ionizable gases** as the propellant. Though the development of electric propulsion was initiated about 30 years ago, the first thrusters were flown in 1980 and 1990 by Europe and the USA (in the former USSR even much earlier) and have been in commercial use since 2000 [4.4.26], [4.4.27]. They are primarily used for orbit control and for north–south station keeping. For interplanetary probes or for missions to comets, electric propulsion is used as the **main propulsion** to achieve a high Δv (e.g., Deep Space 1, Hayabusa, SMART-1, BepiColombo). Because of its toxicity, mercury was replaced by the inert gas xenon early in the development. It is stored in high-pressure tanks, with the high pressure reduced to the final operating pressure and fed to the engines by means of dedicated flow control devices. Most of the electric propulsion systems have, compared to chemical thrusters, a very **low thrust**

level, in the order of some millinewtons, up to about 1 N, but an overall performance level exceeding that of chemical propulsion by a factor of 10 to 20. As a direct consequence, the required amount of propellant is drastically reduced. However, these engines must be qualified for much higher thrust durations and need a relatively high amount of thrust-level-dependent electrical energy, ranging from some hundred watts to several kilowatts, which has to be provided by the satellite or probe.

Figure 4.4.2 (right) presents the schematic of an electric propulsion system.

4.4.2.5 Solid Propulsion

Thrusters using solid propellants played a much reduced role for satellites or probes. During the years 1970 to 1990 their application was limited to **apogee engines**. Thereafter they were replaced by bipropellant thrusters, which allowed more efficient and accurate injection into geostationary orbit because of their reignition capability, enabling apogee injection with three or four consecutive maneuvers.

4.4.2.6 Advantages and Disadvantages

Various parameters can be used to compare different propulsion systems:

- Total mass (dry mass plus propellant and pressurant)
- Electrical power demand,
- Complexity and reliability
- Chemical purity
- Engine thrust levels.

Table 4.4.1 contains the typical thrust levels required for the different propulsion maneuvers.

Table 4.4.1: *Thrust levels.*

Maneuver type	Thrust [N]
Attitude control	<1 to 5
Orbit control	10 to 22
Apogee/kick maneuver	400 to 500
Orbit correction	10 to 500

Depending on the type of propulsion, the following typical thrust levels can be achieved:

- Cold gas: 0.1 to 2.0 N
- Monopropellant: 0.5 to 3000 N
- Bipropellant: 10 to far more than 1000 N
- Solid propulsion: 50 to far more than 1000 N
- Electric propulsion: 0.001 N to several newtons.

While the electrical power demand for chemical propulsion systems is very low, it is very high for electric propulsion systems, which is decisive for the achievable thrust level and the specific impulse.

The decision on the type of propulsion system best suited for a special application is determined by the **specific impulse**, which in turn determines the propellant to be carried on-board the satellite or probe (refer to Section 4.4.3). By means of the basic rocket equation, the required propellant mass can be calculated using the velocity increment (Δv in m/s) needed for the entire mission and the specific impulse of the type of propellant considered:

$$m_p / m_s = 1 - \exp\left(-\Delta v / I_{sp}\right)$$

Figure 4.4.3 presents the ratio of propellant mass m_p to satellite mass m_s.

As an average value, about 65% of the complete satellite mass at launch is stored propellant. The remaining 35% is used for the different satellite subsystems and the payload, while 5 to 6% is required for the propulsion system hardware. As a consequence of the 65% propellant weight share of the total mass, the different propulsion system types can only deliver **maximum velocity increments** as presented in Table 4.4.2.

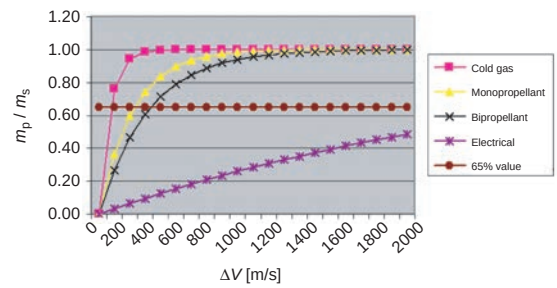


Figure 4.4.3: *Propellant to satellite mass ratio for different propulsion system types.*

Table 4.4.2: *Achievable Δv.*

Propulsion	Δv [m/s]
Cold gas system	74
Monopropellant system	233
Bipropellant system	340
Electrical system	3140

If a higher velocity increment is required for a certain project, it can only be achieved by:

- Reducing the payload.
- Applying new mass-saving technologies.
- Using a higher performance propulsion system; for example, replacing a monopropellant with a bipropellant system.
- Choosing a heavier launcher.

The above considerations do not mean that a bipropellant propulsion system could not provide a "delta-v" of 2360 m/s, such as for a heavy communications satellite. But in such a case the total launch mass of the satellite would amount to 2400 to 2500 kg.

In a number of cases attitude control requires very small impulse bits to perform propellant-efficient corrections around a satellite axis. This minimum **impulse bit** is determined by the thrust level and the minimum firing duration of the thruster. Cold gas and monopropellant thrusters have a clear advantage because of the low thrust levels they are able to generate. But even bipropellant thrusters in the 10 to 22 N range with fast-acting solenoid valves in the order of 2 to 5 ms can fulfill the attitude control requirements for small impulse bits, in particular for heavy communications satellites.

For science satellites with special requirements for **chemical purity**, cold gas systems are preferred for attitude control because contamination of optical instruments or sensors by catalyst particles or unburnt propellant cannot be fully excluded for monopropellant or bipropellant thrusters. Should monopropellant or bipropellant thrusters be mandatory because of the needed velocity increment, the thrusters have to be positioned such that contamination can be minimized or even excluded.

Complexity and **reliability** are other important parameters. The different propulsion systems are discussed in detail in Section 4.4.6. The higher the number of components, the more complex the technology, the more difficulties exist in achieving the required reliability for long and demanding missions. The requirement of high reliability can only be fulfilled with the implementation of redundancies within the system and by extensive development and qualification testing on the component and system levels.

4.4.3 Propellants

Besides the design and optimization of the thrusters, the choice of propellant determines the performance of a propulsion system. As well as the specific impulse, density, freezing and boiling point, toxicity and storability are parameters of importance. The main parameters for the most important propellants used for chemical and electrical propulsion systems are summarized in Table 4.4.3. For satellites with long-duration missions the fuel **storage properties** are of great importance [4.4.4], [4.4.20].

Cold Gas Propellants

The primary choice for cold gas systems is **nitrogen** since it is easier to store compared to helium or hydrogen and less risky with respect to leakage because of its relatively large molecular size. In cases where nitrogen cannot be used because of special requirements of the scientific payload, argon might be a suitable alternative.

Monopropellants

Hydrogen peroxide (H_2O_2) – in high concentration also known as HTP (High Test Peroxide) – was used in the early 1960s as a monopropellant. Structures made of silver or platinum were used as a catalyst, decomposing H_2O_2 into water and oxygen in an exothermic process. The fundamental problem with such propulsion systems was the **inadequate storability** of H_2O_2 for a long period of time. Even under very favorable conditions, H_2O_2 tends to decompose and creates a continuous pressure increase. In spite of provisions such as relief valves etc., this phenomenon led to the total failure of propulsion systems used for the US satellites ATS 1 and Syncom 1.

Hydrazine (N_2H_4) provided the solution for the above problems combined with a performance increase of about 30% compared to H_2O_2. The development of the **Shell 405 catalyst** made the transfer

Table 4.4.3: *Characteristic data of storable propellants and gases.*

Parameter \ Propellant	Nitrogen (N_2)	Hydrogen peroxide (H_2O_2)	Hydrazine (N_2H_4)	Monomethyl hydrazine (MMH)	Unsymmetrical dimethyl-hydrazine (UDMH)	Nitrogen tetroxide (N_2O_4)	Xenon (Xe)
Specific impulse (N s/kg)	700	1700	2300	3200	3200	3200	35 000
Density (kg/dm3)	0.0012 5	1.447	1.008	0.876	0.793	1.447	0.0059
Freezing point (°C)	–210	–0.41	1.55	–52.3	–57.2	–11.2	–111.7
Corrosion/material compatibility	Very stable	Catalytic decomposition	Catalytic decomposition	Catalytic decomposition	Catalytic decomposition	Creates acid with water	Very stable
Toxicity/health risk	–	Acidic	Toxic, carcinogenic	Toxic, carcinogenic	Toxic, carcinogenic	Very toxic, acidic	–
Flammability	–	Explosive self - decomposition	Hypergolic with oxidizer	Hypergolic with oxidizer	Hypergolic with oxidizer	Hypergolic with fuel	–

4

to hydrazine possible. Shell 405 guaranteed even for long-duration missions a reliable and spontaneous exothermic decomposition of hydrazine, as $3N_2H_4 = 4NH_3 + N_2$, followed by an endothermic dissociation of NH_3 into hydrogen and nitrogen. In 1967 and 1968 ATS III and Intelsat III were the first satellites to be launched with hydrazine propulsion systems.

For the two US Viking Mars missions in 1975 a modified hydrazine quality was developed. During tests with the braking engine of the descent stage, hydrogen cyanides were detected, which could have killed any traces of life to be detected during the landing phase. These gases were created by the pyrolysis of aniline in the standard or monopropellant grade hydrazine. Therefore a drastic reduction of the aniline, iron, water and carbon content led to the "high-purity grade" of hydrazine. The reduction of aniline had a further positive effect by largely improving the cold start capability of the thrusters.

Bipropellants

The search for propulsion systems of higher performance was necessitated by satellites with increasingly higher launch masses. Compared to monopropellant hydrazine, bipropellant propulsion systems further reduced the propellant mass by about 25%. From a variety of different propellants, unsymmetrical dimethylhydrazine (**UDMH**) and monomethylhydrazine (**MMH**) were finally selected as fuel and $\mathbf{N_2O_4}$ as oxidizer. In the former USSR, UDMH was the primary choice, while in the western world the thrusters were developed with MMH as fuel. The Symphonie satellites – the very first to use bipropellants – burnt Aerozin 50 (50% hydrazine and 50% UDMH) in their orbit control thrusters and apogee engine. The follow-on propulsion systems used MMH, which was much easier to handle.

To avoid stress corrosion defects in titanium propellant tanks, 1 to 3% of nitrogen oxide (NO) per weight by volume is added to the oxidizer N_2O_4 (NTO = nitrogen tetroxide). This mixture is known as **MON 1** or **MON 3** (MON = Mixed Oxides of Nitrogen) and is used in the 1% or 3% version in almost all bipropellant propulsion systems.

Thrusters for electric propulsion require propellants which can be easily evaporated and ionized and which have a high molecular weight. Therefore in the years 1960 to 1990 **mercury** was the first choice fulfilling most of the above criteria, but it had to be dropped because of its toxicity. Thus the development of thrusters for electric propulsion concentrated on the use of the colorless and odorless inert gas **xenon**, which can be stored in high-pressure gas tanks. Xenon has a high molecular weight and can be quite easily ionized.

Green Propellants

Under the name of "green propellants" investigations are underway to find new propellants for launchers as well as satellite propulsion which are **environmentally**

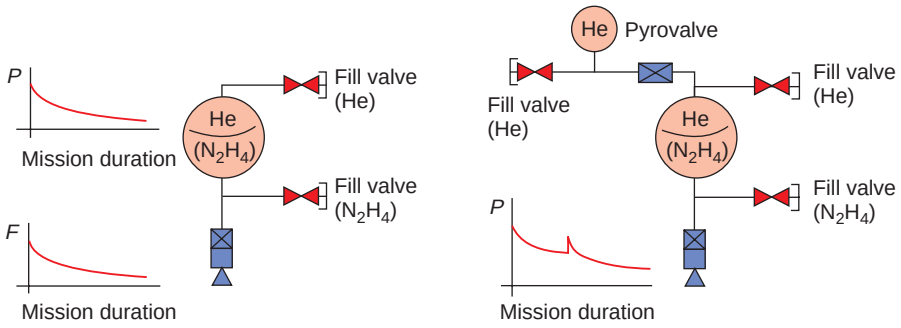


Figure 4.4.4: *Blow-down mode (left) and blow-down mode with repressurization (right).*

friendly and which can be better handled than conventional propellants.

One of the favorites is the above-mentioned hydrogen peroxide (H_2O_2), which, when decomposed, does not generate products polluting the environment. However, hydrogen peroxide or other green propellants such as butane, **water** or even **laughing gas** (N_2O) (systems developed and flown by SSTL, UK) do not deliver performance comparable to that of hydrazine [4.4.21], [4.4.22].

An alternative propellant, which might be equal to or even better than hydrazine, is **ADN (Ammonium Dinitramide)**. Thrusters for ADN have been in development at SSC (Swedish Space Corporation) since 1997. This propellant is decomposed catalytically similar to hydrazine and has a 24% higher density and a 6% higher performance, but at a higher decomposition temperature than hydrazine [4.4.23].

HAN (Hydroxylammonium Nitrate) is also mentioned as a promising candidate with a performance exceeding that of hydrazine. A corresponding thruster development is presently being undertaken in the USA by Aerojet.

Whether the advantages of green propellants are finally sufficient to replace the reliable and efficient hydrazine propulsion systems can only be answered in the future.

The various **materials** used in propulsion systems are subjected to intensive screening and tests regarding their compatibility with the different propellants. Depending on the spacecraft mission, they are in contact with these propellants in the liquid or vapor phase for up to 15 years or more at varying pressures and temperatures. This is valid for metallic materials (keyword: stress corrosion) as well as elastomers (keyword: catalytic/chemical properties). This subject is very complex and cannot be discussed here in any more detail.

4.4.4 Feed Systems and Propellant Storage

All propulsion systems mentioned in Section 4.4.2 store their propellants in tanks, from where they are fed to the thrusters.

In **cold gas systems**, the gas is stored in high-pressure tanks at about 300 bar and is reduced by a pressure regulator downstream of the tank to reduce the pressure to the operating pressure of the cold gas thrusters, normally in the order of 1 to 5 bar.

Unlike in cold gas systems, in **monopropellant or bipropellant systems** the propellant has to be **actively fed** from the tank(s) to the thrusters. In general, two different principles are applied:

- Expulsion in the blow-down mode (thereby reducing pressure)
- Expulsion at constant pressure.

4.4.4.1 Blow-Down Mode

Expulsion that continuously reduces pressure is known as the **blow-down mode**. The tank of the propulsion systems using this method is filled with the propellant (e.g., hydrazine) and the pressurant gas, nitrogen or helium, as presented in Figure 4.4.4 (left).

At the beginning of the mission (BOL) the pressure in the tank corresponds to the maximum **operating pressure** of the thrusters; for hydrazine thrusters

about 22 to 30 bar. Propellant is expelled from the tank as soon as one or more thrusters are operated and the pressure in the tank decreases according to the amount of propellant consumed. Corresponding to the decreasing tank pressure, the engine thrust decreases as well. At the end of the mission (EOL) the remaining pressure in the tank should correspond to the minimum operating pressure of the thrusters; for hydrazine thrusters this is in the order of 5.0 bar. To achieve this, the gas volume should amount to about 25% of the entire tank volume. The thrusters have to be capable of reliable and efficient operation over a pressure range of 4:1, the so-called **blow-down ratio**. During the development of the attitude control algorithms, the continuously changing thrust and impulse bit values have to be considered; thruster performance is of course well known from the relevant acceptance and qualification tests. Blow-down systems are always reasonable when the propellant demand is not too high, otherwise the tank or tanks get too large or too many tanks would be required.

A special case of the blow-down concept is a system with **repressurization** as in Figure 4.4.4 (right). It is advantageous in cases where, at a certain time in the mission, a higher thruster performance (thrust level) is required but is not available because the tank pressure and thus the thrust is reduced to levels not able to support this requirement.

Helium or nitrogen is stored in an additional tank, which is released at a certain time in the mission by opening a pyrotechnic valve. This leads to a pressure increase in the propellant tank, thereby increasing the thrust as well. The mission time for repressurization, and the volume and pressure of the additional tank, have to be determined such that during repressurization the maximum operating pressure of the propellant tank, the thrusters and all the other components of the propulsion system are not exceeded [4.4.1].

4.4.4.2 Expulsion at Constant Pressure

While turbopumps are used to feed the propellants to the engines in launcher propulsion systems, this principle was not accepted for satellite propulsion for several reasons, though corresponding developments existed in the past. One of the main reasons was the problem of **miniaturization** of mechanical elements and their reliability over a period of 15 years.

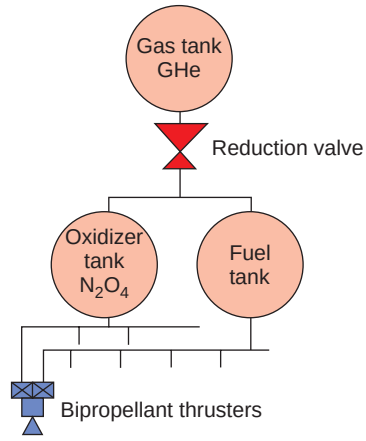


Figure 4.4.5: *Pressure-regulated bipropellant system.*

In a **pressure-regulated** propulsion system the propellant tanks are pressurized with helium or in rare cases with nitrogen out of a separate pressurant tank via a pressure regulator. With this method, the pressure is kept at a constant value between BOL and EOL. This assures a constant propellant supply pressure up to the thrusters over the entire mission. Figure 4.4.5 presents a typical configuration of a pressure-regulated system of the type mainly used in bipropellant propulsion systems.

During the launch campaign the propellant tanks are loaded to 95%. The remaining 5% is filled with pressurant gas to about 4 bar, sufficient to tolerate the maximum allowable pressure increase caused by temperature increases.

After the launch, the propellant tanks are pressurized to the **nominal operating pressure** of about 17 bar. The pressure regulator has to keep the operating pressure constant over the entire mission. This is true for long-duration interplanetary missions but not for commercial communication satellites. The reason for this difference is explained in detail in Section 4.4.6.2.

The goal of pressure regulation is to maintain a constant thrust and constant impulse bits at identical firing conditions. The **specific impulse** is – neglecting aging effects – constant as well. As the engines run close to their optimum operating conditions, the specific impulse is high and the propellant consumption

lower than for operation in blow-down mode. This performance optimization is, however, the result of increased complexity, as demonstrated in Section 4.4.6.2 [4.4.1]. An advantage of the pressure-regulated system is that the propellant tanks are only negligibly larger than required to store the propellant.

4.4.4.3 Propellant Storage

Two main aspects have to be considered when designing a propulsion system: which thrusters are available; and which propellant tanks are available.

New developments should be avoided because of cost. The following section deals with the storage aspects of the propellant.

The storage of gases or the propellant for electric propulsion (xenon) is relatively simple. In general, spherical tanks are used, which are designed for high pressures up to 300 bar. They are mounted in the satellite structure at their spherical poles with one fixed and one flexible bearing to allow for expansion during pressurization. The tanks have only one port connecting the tubing system for loading the gas and its withdrawal during the mission. In the past **all-titanium tanks** were produced from the heat-treated alloy Ti6Al4V. In the meantime so-called **composite tanks** have been more frequently used because of their significant weight advantage [4.4.12], [4.4.18], [4.4.19]. They consist of a thin-walled titanium tank called a liner which is to a high extent leakproof. In the past this liner was wrapped with Kevlar; today Kevlar has been replaced by carbon fiber materials (Figure 4.4.6). After wrapping with the fiber material the tank is soaked in an epoxy resin and subjected to a thermal curing process to achieve the required strength.

The storage of liquid propellant requires completely different technologies compared to the storage of gases, as a zero-*g* environment has a significant influence on the orientation of the propellant.

In **spin-stabilized satellites** with a significantly high rotation speed, the centrifugal force will take care of the propellant orientation, as presented in Figure 4.4.7.

Although only two tank ports would be sufficient during the mission of a spinning satellite, one requirement is not fulfilled by this design, namely the possibility of propellant unloading on the ground because



Figure 4.4.6: *High-pressure composite tanks (carbon fiber or Kevlar) (Source: ASTRIUM).*

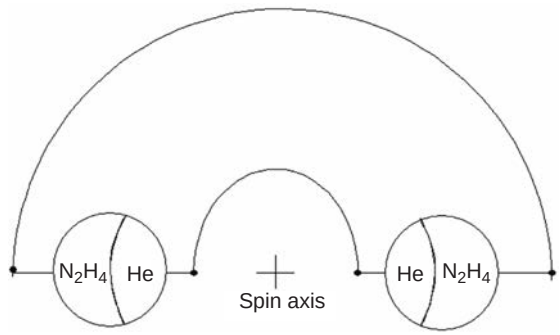


Figure 4.4.7: *Propellant orientation during spin.*



Figure 4.4.8: *Giotto (teardrop) and MSG tank (Source: ATK PSI and ASTRIUM).*

of an aborted launch etc. This requirement can be met by a third propellant port displaced 90° in the equatorial plane of the tank (as realized for the MSG tank, Figure 4.4.8 right) or by a very special tank design, the so-called **teardrop tank** (Figure 4.4.8 left), which was installed in the hydrazine propulsion system of the

Hipparcos satellite and in ESA's Halley Comet probe Giotto. If properly installed in the spacecraft, the lower tank port can be used for propellant withdrawal for the thrusters and for propellant unloading on the ground [4.4.15], [4.4.18].

For **nonspinning satellites or probes** a **propellant orientation method** has to be found which operates reliably under zero-*g* conditions. This method has to be insensitive to disturbing torques and accelerations generated by the operation of the thrusters on-board the spacecraft as well.

One solution for the orientation of propellants or the separation of the propellants and the pressurant gas is to use diaphragms or bladders inside the propellant tank. They can always be used when propellant-compatible materials are available. While this material is available for hydrazine, no suitable and long-time **oxidizer-compatible materials** have been developed.

ATK PSI Operations, USA, is the world's leading company for the development and production of **diaphragm tanks** and has produced several thousand tanks during the last 50 years. On contract to the US Air Force, in the early 1970s TRW developed a hydrazine-compatible material, AF-E-332, replacing the less compatible material EPT 10. Diaphragms produced by ATK PSI for spherical tanks have a hemispherical configuration (Figure 4.4.9). Ribs on the inside improve the expulsion efficiency and a rib at the diaphragm equator is used for fixation in the tank equator.

During final integration of the tank, the diaphragm is installed in the lower hemisphere by means of a clamping ring; thereafter the upper hemisphere is installed and all three parts are connected in one single welding process.

Diaphragm tanks (Figure 4.4.10) have several advantages. Since they can be qualified for up to 100 loading and unloading cycles, they can be easily tested on the component level as well as on subsystem and system levels. They prevent propellant sloshing to a great extent and are capable of orienting the propellant even at high disturbance torques and accelerations. Regardless of whether the tank is filled to 40% or 90% or whether the propellant port is located against Earth's gravity, the tank can be unloaded on the ground and can be launched in this configuration. It can even orient and feed the propellant to the thrusters



Figure 4.4.9: *Different-size diaphragms (Source: ATK PSI Operations) [4.4.18].*

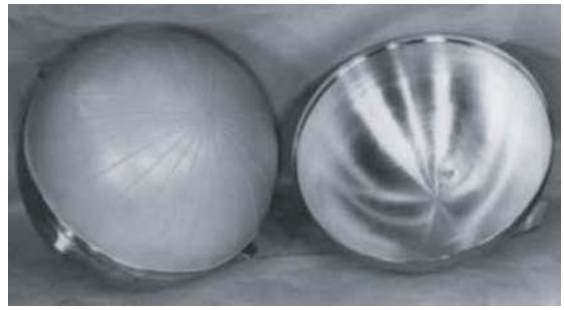


Figure 4.4.10: *Diaphragm tank (Source: ATK PSI).*

under disturbance torques. While the tank is emptied, the diaphragm is reversed until its inner surface comes into contact with the lower tank hemisphere. The propellant between the ribs is forced to the propellant outlet such that an expulsion efficiency of 99.5% can be guaranteed.

But disadvantages exist for tanks with **rubber diaphragms**. In particular, gas penetration and **permeability** have to be mentioned. Helium penetrates the diaphragm material and dissolves in the propellant until helium saturation is achieved. Hydrazine vapors can penetrate the diaphragm in the opposite direction and enter the pressurant part of the propulsion system. Therefore, even those components not directly in contact with the propellant should be hydrazine compatible.

Furthermore, the thermal control subsystem should prevent hydrazine vapors from freezing and thereby endangering or damaging upstream components.

An alternative to the diaphragm tank is the **bladder tank**. The propellant is stored in a cylindrical bladder and the expulsion to the thrusters is achieved through a perforated tube in the center of the tank.

As for the diaphragm tank, the pressurant and the propellant are in the same tank. Again, to achieve a blow-down ratio of 4:1, the gas volume must be in the order of 25% of the total tank volume. If propellant is consumed during the thruster maneuvers, the bladder will fold itself around the central tube and expel the propellant with good expulsion efficiency.

The US company ARDE Inc. has specialized in the development and production of propellant tanks with **metallic diaphragms**. The major reasons for the application of metallic diaphragms are as follows:

- Long-term storage at very low or zero permeation rates.
- Storage of propellants being incompatible with diaphragms or bladders made from rubber.
- Severe requirements concerning sloshing and center of gravity movement.

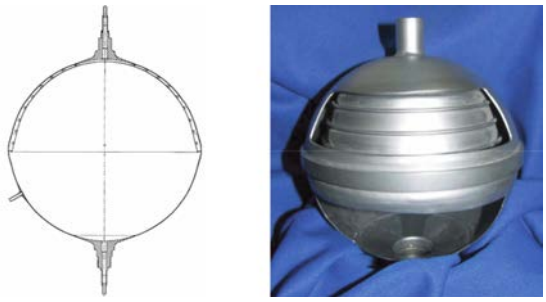


Figure 4.4.11: *Drawing and hardware of a metallic diaphragm tank (Source: ARDE Inc.) [4.4.14].*



Figure 4.4.12: *Partially and completely reversed metallic diaphragm (Source: ARDE Inc.).*

A tank with a metallic diaphragm and, in the partial cut-away, the metallic diaphragm of the tank are presented in Figure 4.4.11.

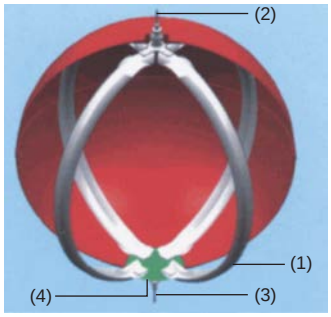
In contrast to the rubber diaphragm, the metallic diaphragm is equipped with concentric ribs which allow a continuous **reversal** of the diaphragm. Two steps of the diaphragm reversal process are presented in Figure 4.4.12. The metallic diaphragm guarantees complete tightness between propellant and pressurant and is compatible with bipropellants as well, even the oxidizer. It can achieve an expulsion efficiency of about 97%. One of the major disadvantages of the metallic diaphragm is the fact that the reversal can be performed just once. The consequence is that acceptance verification of a propulsion subsystem cannot include a diaphragm reversal. Launching with partially filled tanks can cause problems as well.

As a competitive alternative for diaphragm tanks for applications involving monopropellant and bipropellant propulsion systems, **surface tension tanks** have been developed [4.4.13], [4.4.15], [4.4.16], [4.4.17], [4.4.18]. The motivation for the development of this type of tank was the fact that the satellites became larger and heavier and required more efficient propulsion systems. The answer was to use bipropellant propulsion systems, but for the storage of the oxidizer, rubber diaphragm tanks were not suitable and metallic diaphragms never found their way to the market of commercial communications satellites.

Though surface tension forces are rather low, they play an important role under zero-*g* conditions for the orientation of liquids. Surface tension tanks are equipped with propellant management devices (PMDs) capable of continuously transporting the propellant to the tank outlet. These devices make use of surface tension forces in narrow gaps by means of vanes, sponge-like structures or traps and can be refilled depending upon their design.

The design of the propellant management devices depends upon factors such as:

- Type of propellant (hydrazine, MMH, NTO)
- Partial or complete fill level at launch
- Horizontal transport of the launch vehicle on the ground after propellant loading
- The size of the disturbance torques and accelerations and their direction in orbit
- The maximum flow rate during operation of the thrusters
- Point in time and quantity of first propellant withdrawal



(1) Vanes
(2) Pressurization gas inlet
(3) Fuel outlet
(4) Swamp plate

Figure 4.4.13: *Tank with PMD made of vanes (Source: ASTRIUM).*



Figure 4.4.14: *Propellant refillable reservoir (Source: ASTRIUM).*

- Quantity of propellant required for satellite orientation/position recovery in case of loss of orientation.

The Globalstar satellite's orbit and attitude control was performed with 1.0 N thrusters and four thrusters firing simultaneously to generate an acceleration of about $1 \cdot 10^{-3} g$. Its hydrazine tank (Figure 4.4.13) was equipped with a PMD of four symmetric vanes only.

The higher the requirements for the tank, the more complex are the propellant management devices.

Simple vanes are not at all sufficient for **high flow rates** and disturbance torques as well as accelerations in different directions. Launchers like Soyuz or Proton are **transported horizontally** to the launch site with the satellite installed in the upper stage and fully loaded with propellant. This was a new requirement



Figure 4.4.15: *Astrium's surface tension tank OST 22/X (Source: ASTRIUM).*

for the PMD and difficult to implement. A propellant refillable reservoir (PRR) has to be installed next to the propellant outlet port in the tank, the contents of which must be sufficient for carrying out the initial satellite maneuvers. Any refilling of the PRR must take place during a phase of low- or no disturbance accelerations thereafter.

Figure 4.4.14 presents the hardware of a PRR able to fulfill all the requirements of a state-of the-art tank.

From a strength point of view, the optimum tank configuration is that of a sphere mounted in the satellite structure at the poles or in its equatorial plane. In order to enable the Ariane 5 launcher to launch two satellites at a time, the height of the satellite has to be as low as possible. Therefore the length of the propellant tanks should be as low as possible as well, being often the dominating component in a satellite. This can be achieved by not shaping the tank upper and lower domes hemispherically. The tank presented in Figure 4.4.15 is a good example of a tank with a "Cassini"-type half shell.

Using cylindrical center sections of different heights this tank can be adapted to different volumes and can be qualified for propellant volumes between 700 and 1450 dm^3.

4.4.5 Cold Gas Propulsion Systems

Cold gas systems were used for attitude control of the first satellites because of their rather simple design and high reliability. At present they are only used if:

- The required **total impulse** is low.
- The required **pointing accuracy** is very high.
- Chemical propellants are prohibited because of **chemical purity** requirements for sensors or payload.

The ESA satellite Hipparcos (because of the high pointing accuracy of its astrometrical payload), ASTROSPAS, CHAMP and GRACE, as well as EURECA (because of the microgravity experiments) and the early "Symphonie" communications satellites, used cold gas systems for their attitude control maneuvers.

The propellants, mostly nitrogen, are stored in one or more all-titanium or composite tanks at a pressure of up to 300 bar. However, not only the availability of a suitable tank, but all the other components in the high-pressure section of the propulsion system, like valves or pressure reduction devices, determine the maximum operating pressure.

The complete flow schematic of a cold gas system is presented in Figure 4.4.16. At least one **temperature sensor** is fixed on the tank shell with the purpose of:

- **Measuring the tank temperature** during gas loading to avoid exceeding the maximum tank temperature during pressure increase.
- **Monitoring the temperature** during the mission, in order to determine the remaining quantity of gas in the tank using the pressure–volume–temperature (PVT) method.

The loading and, if necessary, unloading is performed by means of the fill and vent valve; pressure monitoring is performed by the high-pressure transducer. A filter in the high-pressure section protects the pressure regulator from particle contamination. In addition, most of the present pressure regulators are equipped

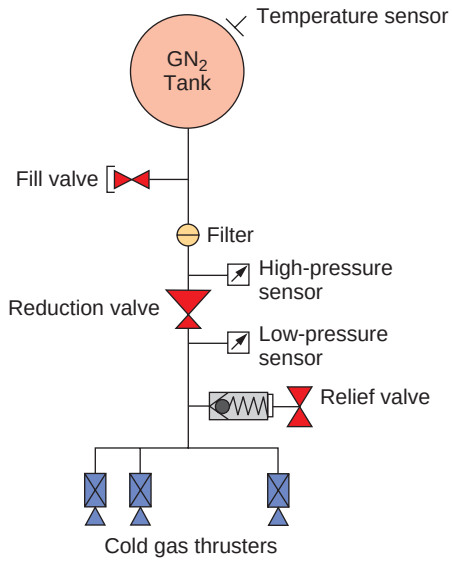


Figure 4.4.16: *Flow schematic of a cold gas propulsion system.*

with their own inlet filter for protection during acceptance testing at the component level. The pressure regulator reduces the high pressure – 300 bar, for example – to the operating pressure of the cold gas thrusters, 1 to 5 bar. The required thrust level more or less determines the operating pressure in the low-pressure section, as the thrust is proportional to the inlet pressure into the cold gas thruster. The cold gas thrusters of the Hipparcos satellite delivered a thrust of 20 mN at an operating pressure of 1.5 bar.

A **low-pressure transducer** is located downstream of the pressure regulator to monitor the inlet pressure of the thrusters. This is important, if the regulated pressure is not constant; its characteristics may depend on the inlet pressure. The low-pressure section is protected against a failure of the pressure regulator by a relief valve which will open at a certain "delta-p" (about 2 to 3 bar) above the nominal operating pressure and vent the gas. During the mission a burst disk at the relief valve outlet (with a burst pressure 1 or 2 bar above the cracking pressure of the relief valve) protects the system against a potential leak of this valve. Should this really occur, small nozzles with an offset of 180° will vent the gas without creating disturbance torques in the satellite. In the meantime pressure regulators exist with two almost identical regulator stages in

series, the lower regulator stage having a regulating pressure about 0.5 bar higher than the first stage. Due to this design feature, the total loss of a regulator is very unlikely. Therefore the relief valve, burst disk and nozzle arrangement could be avoided, if ground safety considerations did not mandate these devices.

Cold gas systems are still used, as proven by satellites such as CHAMP and GRACE. Besides its hydrazine propulsion system, the German satellite TanDEM-X uses a cold gas system to keep in constellation with its partner satellite, TerraSAR-X.

4.4.5.1 Cold Gas Thruster

As mentioned previously, cold gas thrusters consist of a solenoid valve and a nozzle and are the simplest thruster design.

The largest part of the thruster is the **solenoid valve**. As shown in Figure 4.4.17 (on the right), the valve is protected against particle contamination at its inlet by a conical filter.

In the cylindrical part of the drawing in Figure 4.4.17 (on the left) a **conical nozzle** (hatched) can be recognized. To achieve a thrust of 20 mN at an operating pressure of 1.5 bar, the throat diameter of the nozzle must amount to less than 0.3 mm.

As soon as the valve receives the electrical command to open, a magnetic field is established which leads to a movement of the pintle. At a stroke of about 0.3 mm, the gas passes the seal/seat area, leaves the nozzle and, depending on the operational mode, creates a continuous thrust or an **impulse bit**. As soon as the electrical activation is terminated, the valve is closed by a spring. The force of the spring acts as a closing force in the nonoperating condition to prevent the valve from leaking, pressing against a rubber seat/seal configuration.

Solenoid valves have to be able to open and close within a few milliseconds, if very small impulse bits are to be achieved. Opening and closing response times in the order of 10 milliseconds and a thrust level of 20 millinewtons make minimum impulse bits of $2 \cdot 10^{-4}$ N s possible. High-precision attitude control requires the qualification of the thrusters for several million pulses and many hours of continuous operation, such as for Hipparcos, which had to be operated safely during the entire multiyear mission.

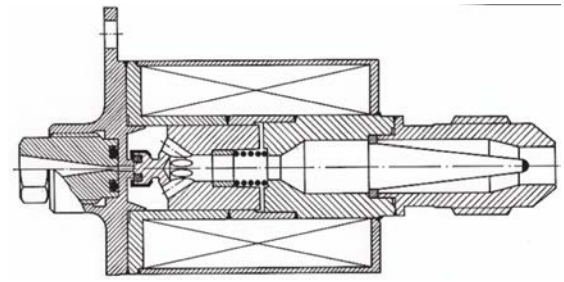


Figure 4.4.17: *Cross-section of a cold gas thruster.*



Figure 4.4.18: *Cluster with four cold gas thrusters of the ASTRO-SPAS platform (Source: ASTRIUM).*

These cold gas thrusters have been used for attitude control of the EURECA and ASTRO-SPAS platforms (Figure 4.4.18) as well as for Hipparcos.

Cold gas thrusters with much higher thrust levels are used for roll control of the upper stages of launchers as well, conveniently using as propellants the gas from the propellant tanks as the operating medium.

A special form of cold gas thrusters is known as a **resistojet**. Before leaving the nozzle, the gas is heated by electrical resistance heater elements. This increases the enthalpy of the gas, leading to a higher exhaust velocity, thereby increasing the specific impulse in the order of 35 to 40%, and in turn reducing the propellant demand by this percentage. A prerequisite for this increase in performance is the availability of electrical power. The resistojet is described in Section 4.4.6.1 as a special design of a hydrazine thruster, the "electrothermal hydrazine thruster."

Table 4.4.4: *Characteristics of different gases.*

Gas	M [g/mol]	k	I_{sp} [N s/kg]
Hydrogen	2	1.4	290
Helium	4	1.659	176
Nitrogen	28	1.4	70
Freon 14	88	1.22	49
Argon	40	1.667	50
N_2O (laughing gas)	44	1.27	67

4.4.5.2 Design Aspects

Operating media for cold gas thrusters are gases such as helium, nitrogen, Freon, hydrogen and argon. With the exception of hydrogen, all of the above gases have been used. Table 4.4.4 presents their characteristic values.

Looking exclusively at the specific impulse, hydrogen would be the ideal operating medium. However, the low density requiring extremely large tanks disqualifies hydrogen.

Helium as well is not the best choice for long mission durations due to the risk of leakage. Nitrogen therefore dominates in cold gas systems, because it is nontoxic, environmentally friendly and easy to handle.

On the basis of a known velocity increment Δv and the maximum satellite launch mass m_s, the necessary **propellant mass** m_p can be determined using the rocket equation [4.4.1], [4.4.2], [4.4.4]:

$$m_p = m_s \cdot \left[1 - \exp\left(-\Delta v / I_{sp}\right)\right] \qquad (4.4.1)$$

with m_p and m_s in kg, Δv in m/s and I_{sp} in N s/kg or m/s.

Should the total impulse required be given, the necessary propellant mass is calculated as follows:

$$m_p = I_{tot} / I_{sp}$$

with I_{tot} in N s and I_{sp} in N s/kg.

Margins have to be added to the calculated propellant mass to compensate for leakage losses, loading uncertainties and additional margins required by the spacecraft specification.

The calculation of the tank volume is performed using the general gas equation:

$$P \cdot V_T = m_G \cdot R \cdot T \cdot Z \qquad (4.4.2)$$

where:
P = max. tank pressure [N/m^2],
V_T = tank volume [m^3],
R = gas constant [m^2/(s^2 K)],
T = gas temperature [K],
Z = compressibility factor [K].

The compressibility Z should not be neglected for nitrogen as it amounts to 0.9954 at 20 bar and 293 K, while it amounts to 1.1182 at 280 bar and 293 K, corresponding to a volume increase of about 11%.

Using the volume V_T the tank dimensions can be derived depending upon the tank configuration selected.

4.4.6 Chemical Propulsion

In the following, "chemical propulsion" refers to "monopropellant" and "bipropellant" systems, though cold gas propulsion belongs in this category as well.

4.4.6.1 Monopropellant Propulsion

As described by "monopropellant," this type of propulsion needs only one propellant to generate thrust. A catalytically or thermally induced exothermal decomposition process produces hot gases, which are expanded through a nozzle. The only propellants which have been used in flight up to now are hydrogen peroxide and hydrazine. The major characteristics of these propellants are summarized in Table 4.4.5.

H_2O_2

Due to the availability of a suitable **catalyst** for the decomposition of hydrogen peroxide (H_2O_2) on the

Table 4.4.5: *Characteristics of monopropellants.*

Monopropellant	Density [g/cm^3]	Specific impulse [N s/kg]
Hydrazine	1.008	2250
Hydrogen peroxide	1.45	1600

basis of a silver or platinum matrix – not available for hydrazine in the 1960s – the first satellites with monopropellant propulsion systems used this type of propellant. The US satellites Syncom and ATS 1 were equipped with hydrogen peroxide monopropellant systems. The major disadvantage of hydrogen peroxide is that it tends to **self-decomposition** even under the best storage conditions. In addition, this propellant did not exist in today's **purity** and **high concentration**. Therefore, the former propulsion systems had to be equipped with relief valves, as described in Section 4.4.5. These relief valves proved to be very unreliable and several satellites lost their propulsion systems due to failures of those valves. Of course one advantage of H_2O_2 has to be mentioned. It belongs to the group of "green propellants" as it is nontoxic and decomposes into hot water vapor (at about 600 °C) and oxygen.

4

Anhydrous Hydrazine

Immediately after the successful development of the Shell 405 catalyst for the decomposition of hydrazine, this propellant became the favorite choice for monopropellant propulsion systems. Hydrazine brought additional advantages:

- Relatively **easy handling**
- Low **self-decomposition rate**
- Good **material compatibility**
- A 40% increase of the **specific impulse** compared to H_2O_2.

Hydrazine decomposes according to the following reactions:

$3N_2H_4 \rightarrow 4NH_3 + N_2$ (exothermal process)
$4NH_3 \rightarrow 2N_2 + 6H_2$ (endothermal process).

Due to the dissociation of ammonia (NH_3) and the resulting loss of energy, this latter process must be minimized for hydrazine thrusters and has to be considered in the design and dimensioning of the catalyst bed. A 100% dissociation of the ammonia would decrease the specific impulse by about 18%.

Catalysts for Hydrazine Decomposition

The breakthrough of hydrazine as a monopropellant is based on the successful development of the Shell 405 catalyst by the Shell Company in 1961 [4.4.11]. The major requirement was an almost unlimited spontaneous restart capability, high performance, and good mechanical and thermal stability.

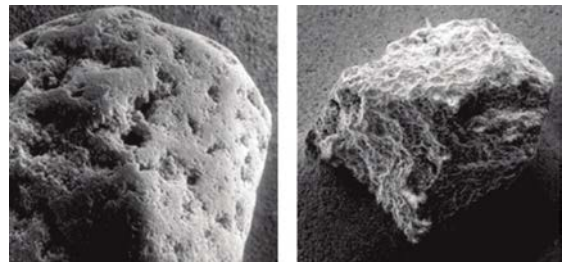


Figure 4.4.19: *Catalyst grains Shell 405 (left) and KC 12 GA (right) (Source: ASTRIUM).*

All presently applied catalysts use an **Al_2O_3 granular material** as carrier and **iridium** as catalyst. Most important is the distribution of the micro- and macropores of the carrier itself. The carrier is delivered in different grain sizes and is subjected to a mechanical rounding process to improve the abrasion resistance of the carrier and to allow better packing into the catalyst bed. The final catalyst is achieved by multiple impregnation steps in an iridium chloride solution. Each impregnation step is followed by a drying process. The number of impregnation steps is determined by the quantity of iridium (33 to 35% by weight) to be used for coating the catalyst grains. The final step in the manufacturing process is reduction in a hydrogen atmosphere at a given temperature. To be able to handle the highly active catalyst, it has to be carefully surface oxidized. Since 2003 the US catalyst for all space applications has been manufactured by the US company Aerojet.

Figure 4.4.19 shows different catalyst grains in a 100-fold magnification as seen by a scanning electron microscope.

To reduce the dependency of the European production of hydrazine thrusters on US products, the GFW (today DLR) decided in 1972 to initiate the development and production of a German catalyst [4.4.9]. The following companies and institutes participated in this development:

- Kali Chemie in Hanover
- ERNO Raumfahrttechnik GmbH in Bremen (today ASTRIUM GmbH ST)
- The DLR at Trauen.

The Kali Chemie Company developed the catalyst with the designation **KC 12 GA**. It was subjected to various tests by DLR-Trauen to investigate and verify its reactivity and stability. The final qualification of the catalyst in three different thrusters of 22, 0.5 and 2.0 N thrust was performed by ERNO.

Today this catalyst is known as **H-KC 12 GA** and is produced by the German company W.C. Heraeus of Hanau. This company also produces a ruthenium/iridium catalyst for space applications. This means that all types of space-qualified catalysts used in Europe are produced by W.C. Heraeus.

Besides the two catalysts mentioned above, CNES and ESRO (today ESA) initiated the development of the so-called CNESRO catalyst, which was exclusively used in thrusters produced by the SEP Company (today Snecma Moteurs). Neither the catalyst nor the thruster is manufactured any more.

This section discussing the properties of catalysts should not be concluded without mentioning the different **damaging mechanisms**. Damage means the reduction or in the worst case the complete loss of catalytic activity. Whether the following damage will really occur depends to a great deal on the quality of the design of the catalyst bed(s), on the catalyst loading process and on the operating modes and conditions during the mission:

Damage can be caused by:

- Generation of voids in the catalyst bed during acceptance and launch vibration
- Loss of iridium on the active surface
- Breakup of catalyst grains due to high thermal stress
- Compacting of the catalyst.

Hydrazine Thrusters

Using the cross-section schematic shown in Figure 4.4.20, the function of a hydrazine thruster is explained in the following [4.4.5], [4.4.6], [4.4.7].

At the beginning one remark should be made. Contrary to bipropellant thrusters, where a chemical reaction between two propellant components takes place – oxidizer and fuel – monopropellant thrusters do not have a **combustion chamber** but rather a **decomposition chamber**.

Upon opening the solenoid valve – the flow control valve (not shown in Figure 4.4.20) – the propellant

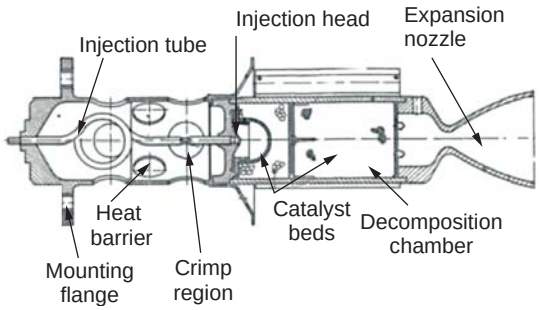


Figure 4.4.20: *Cross-section of a 2.0 N hydrazine thruster (without valve) (Source: ASTRIUM).*

hydrazine passes the **capillary tube** into the **injection head**. For engines with a very low thrust level, this capillary tube has a very low internal diameter of 0.1 to 0.2 mm and is bent in an S-shape or even a complete coil. This supports the flow calibration and gives the tube a certain flexibility to overcome stresses induced by the severe temperature gradients between the inlet side and the injector head. The capillary tube is fixed to the heat barrier on the left and to the injection head on the right by high-temperature brazing or welding in vacuum.

The complete unit consisting of heat barrier, capillary tube and injection head is subjected to a flow calibration in order to achieve a thrust level tolerance of ±3%. A possible method is to reduce the cross-section of the capillary tube in the crimp area by means of a special tool. Simultaneously the flow rate is measured and the process is completed as soon as the predefined flow rate has been achieved.

The configuration of the injection head of the 2.0 N thruster in Figure 4.4.20 looks like a showerhead with several injection bores and forms an integral part of the injector plate. These bores are equally spaced, so that when the injected propellant hits the front of the catalyst bed it is as uniformly distributed as possible. Injection head and catalyst bed are separated by a hemispherical screen. This screen and the appropriate location of the bores prevent catalyst fines from entering and blocking one or more of these bores.

As can be seen in Figures 4.4.20 and 4.4.21, the thruster has two **catalyst beds** of different grain sizes, which again are separated by an appropriately sized screen. The grains in the first catalyst bed where the

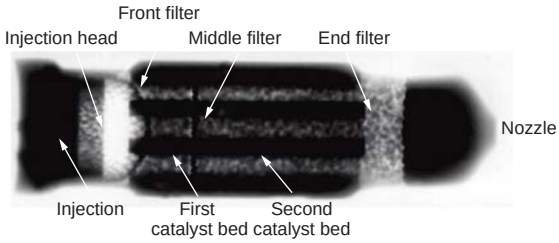


Figure 4.4.21: *X-ray photograph of a hydrazine thruster (Source: ASTRIUM).*

decomposition of the liquid or partially evaporated hydrazine takes place are slightly larger, in the order of < 1.0 mm, while the grains in the second bed are smaller, < 0.7 mm. In the second catalyst bed the decomposition of the hydrazine is completed, but the partial dissociation of ammonia also takes place. For the dimensioning of the catalyst beds a possible degradation has to be considered, in particular the reduction of the active surface of the upstream bed during the mission. The decomposition front will move toward the middle screen. The length of the downstream bed can therefore not only be sized on the basis of a minimization of the ammonia dissociation (as short as possible), but also take the functional lifetime and the mission duration into account.

A characteristic value in the design of the catalyst beds is the **bed loading**, which is in general determined by the maximum flow rate divided by the cross-section of the decomposition chamber. Typical values are between 0.015 and 0.060 g/(s mm^2).

During thruster operation, the maximum decomposition chamber temperature can be in the order of 950 °C. The decomposition products expand through the bell-shaped nozzle depending on the operational case, the steady-state or pulse mode.

The thrusters presented in Figure 4.4.22 are ready for flight, equipped with a flow control valve, catalyst bed heaters and a thermocouple. The thruster on the right is provided with a thermal insulation wrapping to minimize heat loss during the heating up of the catalyst bed and to protect the environment of the thruster from the heat generated during thruster hot firing [4.4.3], [4.4.6].

The **catalyst bed heaters** keep the catalyst at a start-up temperature between 150 and 200 °C prior to firing, to avoid extreme cold starts with high

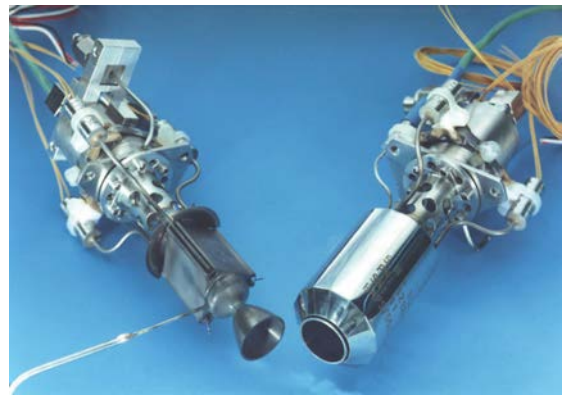


Figure 4.4.22: *Two 5.0 N hydrazine thrusters (Source: ASTRIUM).*

temperature gradients and to prevent the catalyst from damage. However, the thrusters can operate from ambient temperatures and have been qualified – depending on mission requirements – to the required quantity of cold starts.

Electrothermal Hydrazine Thrusters and Arcjets

The efforts to reduce the propellant mass led to thrusters for which the enthalpy of the decomposition products was increased by raising the temperature of the exhaust gases. Higher exhaust gas temperature means increased exhaust velocity, resulting in an improved specific impulse.

The US company Aerojet [4.4.8], [4.4.10], [4.4.24] was very successful in developing **electrothermal hydrazine thrusters** and **hydrazine arcjets.** Figure 4.4.23 shows the electrothermal hydrazine thruster MR-501B, which is similar to the above-mentioned resistojet of cold gas propulsion.

The thruster consists of a **hydrazine gas generator** (in Figure 4.4.23, left) and a chamber with an integrated **resistance heater** and the **exhaust nozzle** (in Figure 4.4.23, right). The gas generator decomposes the hydrazine catalytically in the same way as in the hydrazine thruster. Instead of directly leaving the thruster, the decomposition products are heated up by the resistance heater in the second chamber to a temperature of about 2000 K before they reach the nozzle.

The thruster delivers a thrust of 0.37 to 0.18 N at a blow-down ratio of 3.5:1. The achieved specific



Figure 4.4.23: *Electrothermal hydrazine thruster (Source: Aerojet).*



Figure 4.4.24: *Low-power hydrazine arcjet (Source: Aerojet).*

impulse is quite remarkable. With an electrical power input of about 500 W, the thruster delivers an I_{sp} of 2940 N s/kg.

For the operation of the thruster, **electronic circuitry** is required to limit the current drawn by the resistance heater to 30 A, until the operating temperature is reached.

Figure 4.4.24 presents the Arcjet MR-510 including electronics and power cable.

The design of the **arcjet thruster** is similar to the design of the electrothermal hydrazine thruster: a gas generator is used for this system as well to produce the decomposition products. The heating of the gases is not performed by a resistance heater but in the **light arc** of an arcjet. An electrical power of 1 to 2 kW is necessary to produce a thrust level of 200 to 300 mN. A satellite providing this amount of power can take advantage of thrusters with a specific impulse up to 6000 N s/kg.

The development of an arcjet thruster needs skills in the design of standard hydrazine thrusters and power electronics. Because of the high temperatures in the arcjet – in the arc the temperature reaches 10 000 K – the aspects of material selection and thermal control are of utmost importance.

Hydrazine Propulsion Systems

Monopropellant propulsion systems with hydrazine as propellant have been used with increasing success since the 1970s. Depending on the scientific or commercial application and the mission lifetime, the implementation of redundancies leads to systems of quite different complexity.

The first commercial application of the Shell 405 catalyst took place for the 22 N thrusters of the **spin-stabilized Intelsat III satellite**. This satellite was developed and produced by TRW (today Northrop Grumman). The German company ERNO Raumfahrttechnik GmbH (today Astrium GmbH) delivered two complete hydrazine propulsion systems (Figure 4.4.25).

Despite **redundancy** (two identical half-systems) the Intelsat III propulsion system is not very complex. Each of the two half-systems comprises an axial and a radial thruster, a fill and a drain valve for loading or unloading hydrazine, a pressure transducer and a filter. The two half-systems are interconnected by a normally closed pyrotechnic valve, separating them during a nominal mission.

If a thruster fails the redundant half-system can still fulfill the mission. By opening the pyrotechnic valve the propellant from the half-system with the faulty thruster can be used by the redundant half-system. This of course is only possible if the faulty thruster is not leaking.

Figure 4.4.26 presents the flow schematic of the propulsion system of the ECS and MARECS satellites, three-axis-stabilized communications or weather satellites.

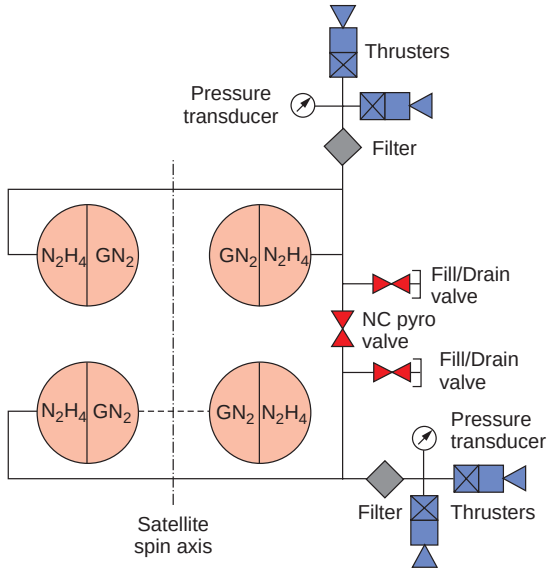


Figure 4.4.25: *Flow schematic of the Intelsat III hydrazine propulsion system (Source: ASTRIUM).*

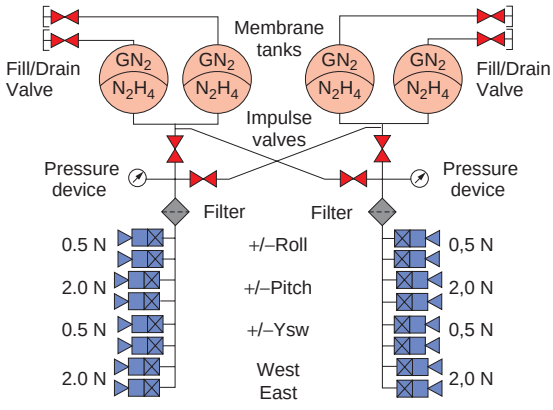


Figure 4.4.26: *Flow schematic of the ECS/MARECS hydrazine propulsion system (Source: ASTRIUM).*

Again two half-systems in a so-called H-configuration provide system redundancy. Switching the latching valves allows propellant to flow from each tank pair to each thruster branch or can isolate a thruster branch in case of leakage.

The propulsion system of the Telecom I satellite (Figure 4.4.27) is similar to that of ECS with the exception of larger propellant tanks.

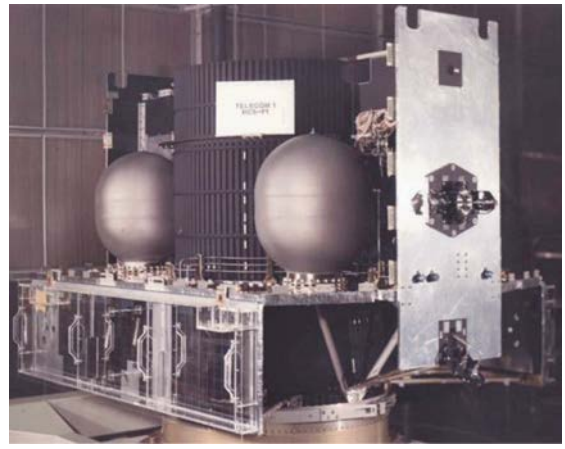


Figure 4.4.27: *Hydrazine propulsion system of Telecom 1 (flight model) (Source: ASTRIUM).*

Due to its low complexity and high reliability, the hydrazine system is also used for roll and attitude control of launchers, such as for the Atlas II, Titan IV, Delta III, H-II A and Ariane 5 rockets, but with thrusters of higher thrust level. The Ariane 5 SCA (Système de Contrôle d'Attitude) uses six to eight hydrazine thrusters with a thrust level of 400 N.

4.4.6.2 Bipropellant Propulsion Systems

Looking at chemical propulsion systems, bipropellant propulsion delivers the highest performance [4.4.1], [4.4.3]. This is true for the **propellant combination** LH_2/LOX, as described in Section 3.3. For satellite propulsion, the propellants MMH and N_2O_4 are best suited due to their excellent storability and good performance. The propellant characteristics are summarized in Table 4.4.3**.** MMH reacts hypergolically and ignites spontaneously once in contact with N_2O_4; no special igniters are necessary. This attribute gives bipropellant engines the capability of pulse mode operation with very short pulse widths substantially lower than 10 milliseconds.

Bipropellant Thrusters

While for monopropellant thrusters the layout of the catalyst bed and the catalyst loading are of major importance, the challenge for bipropellant thrusters is the design of the **injector** and the combustion

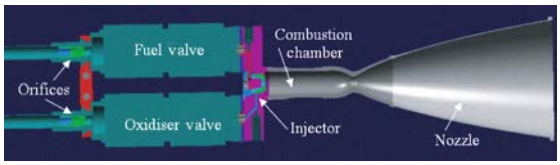


Figure 4.4.28: *Components of a bipropellant thruster (Source: ASTRIUM).*



Figure 4.4.29: *Showerhead (left) and swirl injector (right) (Source: ASTRIUM).*

chamber (Figure 4.4.28). This design determines the performance and stability of the thruster.

Various injector configurations have been designed, developed and tested, such as doublet, self-impinging, triplet, etc.

Figure 4.4.29 makes it clear that bores and channels have to be produced with the highest precision, in order to fulfill the extreme requirements for reproducibility and performance. Furthermore, major factors influencing performance, stability and service life of the thruster are adherence to the mixture ratio, efficient mixing of oxidizer and fuel, and cooling of the injection head, combustion chamber walls and nozzle throat.

The adiabatic flame temperature for MMH and N_2O_4 exceeds 3000 K. Therefore the cooling of the thruster is of great importance. **Cooling methods** applied in thrusters used for satellite propulsion are:

- Film cooling
- Radiation cooling
- Regenerative cooling.

The last method is no longer customary. The present generation of thrusters uses a combination of film and radiation cooling methods.

Because of the high temperatures and the **aggressive propellants**, above all of the oxidizer, the selection of appropriate combustion chamber materials has a big impact on thruster performance and life. The following processes and materials are used for protection of the inner wall of the combustion chamber against oxidation:

(a) Columbium with a complete silicide coating
(b) Platinum without any coating
(c) Rhenium with an iridium coating produced by chemical vapor deposition (CVD).

Evaluating the above processes or materials, the following can be stated:

- Material and processing according to (a) have the lowest potential but have been in use for more than 50 years.
- The material according to (b) is the most reliable with a maximum thermal capability of 1600 K without any coating.
- The material and processing according to (c) have the highest potential with a maximum thermal capability of more than 2000 K. An increase in the specific impulse of over 5% is possible.

For the propellant combination N_2O_4/MMH the optimum mixture ratio with the highest specific impulse amounts to 1.65. A very accurate individual calibration of the oxidizer and fuel flow rates by means of orifices, as shown in Figure 4.4.28, is of utmost importance. Due to the different densities of N_2O_4 and MMH, this mixture ratio has the advantage of resulting in identical tank sizes for oxidizer and fuel, thus avoiding the development of different tank sizes.

The different functions of satellite propulsion systems were presented in Section 4.4.1.1. Present systems are designed for orbit and attitude control and apogee injection. The **thrust levels** for these tasks are quite different. Apogee injection is performed with 400 to 600 N thrusters, while orbit and attitude control make use of 1 to 22 N thrusters. The thruster presented in Figure 4.4.30 has been designed for the propellant combination MON/MMH.

The **combustion chamber** and **nozzle throat** are manufactured using a platinum alloy and are therefore highly temperature resistant and insensitive to oxidation. The **specific impulse amounts** to 3150 N s/kg. The trend toward high capacity and continuously heavier communications satellites is unbroken and propellant demands primarily for the apogee injection maneuvers are very high. Therefore higher performance



Figure 4.4.30: *The 400 N apogee engine (Source: ASTRIUM).*

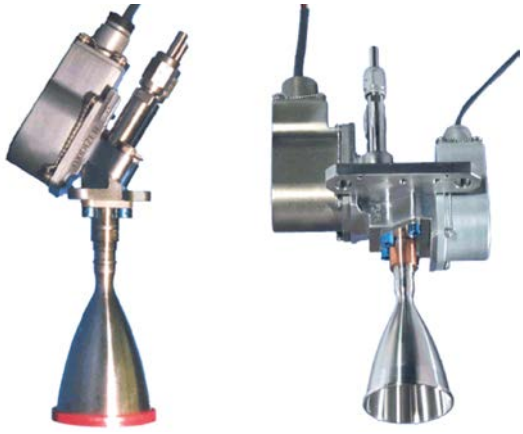


Figure 4.4.31: *The 10 N thruster for orbit and attitude control (single seat left, dual seat right) (Source: ASTRIUM).*

apogee engines are in development in anticipation that an increase of 20 to 30 N s/kg in specific impulse will lead to a propellant mass reduction of 15 to 20 kg.

It is obvious that the performance increase achieved by the application of iridium-coated rhenium or other high-temperature resistant materials such as carbon fibers is of great importance.

Apogee injection is in general performed in three steps/maneuvers. Thereafter the apogee engine has completed its mission. For interplanetary probes this can be completely different with mission durations up

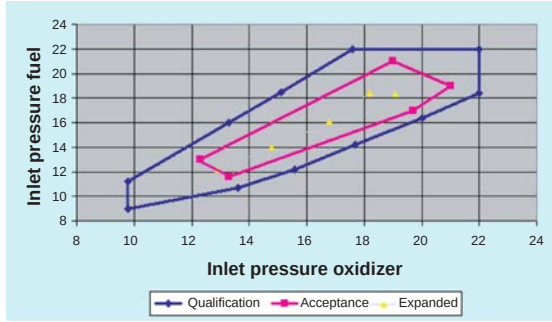


Figure 4.4.32: *Operational range of a 10 N bipropellant thruster (Source: ASTRIUM).*

to 10 or 15 years. Trajectory correction maneuvers or injection into planetary orbits may occur quite late in the mission.

The thruster presented in Figure 4.4.31 for orbit and attitude control is equipped with a platinum combustion chamber. Delivering a vacuum thrust of 10 N at a specific impulse of 2855 N s/kg it belongs to the smallest high-performance thrusters of its kind. It is equipped with two **valves** as is more or less standard today. The upstream valve is a bistable latching valve which can isolate the thruster from the rest of the propulsion system either on the ground or during any phase of the mission. The downstream valve is a monostable solenoid valve which controls the propellant flow to the thruster in the steady-state or pulse mode. In the present case both valves are integrated into one single body (see Figure 4.4.31, right).

A typical **operational range** for oxidizer and fuel inlet pressures is presented in Figure 4.4.32.

For the qualification of a 10 N thruster an extensive test program has to be performed in the steady-state mode (up to 60 hours of continuous firing) and in pulse mode operations (up to 10^6 pulses with various on/off times including negative pulses) thereby verifying the diagram of Figure 4.4.32.

Bipropellant Propulsion Systems

The German/French Symphonie communications satellite was the first to be equipped with a bipropellant propulsion system [4.4.3] (see, e.g., Figure 4.4.33). However, the apogee engine and the orbit control thrusters were still components of two independent propulsion systems. A substantial increase in efficiency and thus cost reduction was connected with the invention

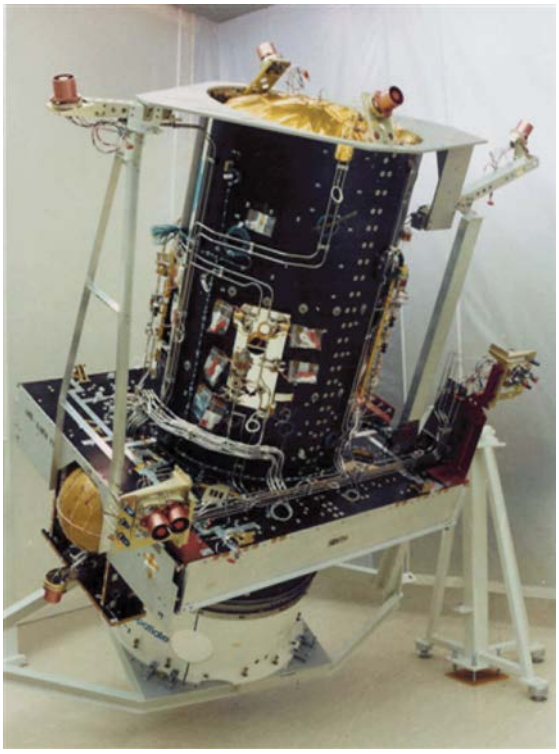


Figure 4.4.33: *Bipropellant propulsion system integrated in the satellite structure (Source: ASTRIUM).*

of the **unified propulsion system** (UPS). It was first designed and produced for the German/French communication satellites TV-SAT and TDF. The word "unified" is explained by the fact that the apogee engine and the orbit and attitude control thrusters get their propellant from the same tank system.

Bipropellant systems (see Figure 4.4.34) with more than 2000 kg of propellant designed for today's communications satellites use about 80% for the apogee injection. In the transfer phase, that is prior to and during the apogee maneuvers, they are **pressure regulated**, taking the helium pressurant from one or more gas tanks (helium tank).

Pressure regulation to achieve the operating pressure of the thrusters is performed by a pressure regulator (PR1). State-of-the-art pressure regulators have built-in redundancy (see Figure 4.4.42) and consist of two units in series. During launch the pressure regulator is protected by a normally closed pyrotechnic valve (PV1). Downstream of the tanks are two normally closed pyrotechnic valves (PV4/5) which act on the ground as mechanical inhibitors for safety reasons, as required by most launch authorities. After pressurization of the tanks and opening of all normally closed pyrotechnic valves, the propulsion system is ready for operation. The first major maneuvers are for apogee injection. Due to the pressure-regulated mode the operational conditions for all apogee maneuvers are identical. As soon as the satellite has reached its geostationary orbit, the two normally open pyrotechnic valves (PV2/3) are closed and the complete pressure supply system is isolated from the remaining propulsion system. As a consequence all subsequent orbit and attitude control maneuvers are performed in a blow-down mode. As a large quantity of propellant has been withdrawn from the propellant tanks because of the apogee injection and are now filled with helium, the blow-down ratio is quite low, normally less than 2:1.

The thrusters of the propulsion system shown in the schematic of Figure 4.4.34 are equipped with monostable solenoid valves only. Therefore the two bistable latching valves (BLV1/2) installed in front of each thruster branch act as a third barrier, together with the two normally closed pyrotechnic valves (PV4/5) for ground safety. Should the thrusters be equipped with dual-seat valves as shown in Figure 4.4.31 (right), these latching valves can be omitted.

The two propellant tanks of the propulsion system presented in Figure 4.4.33, with tank shapes as depicted in Figure 4.4.15, are integrated in the central compartment of the satellite. The 10 N thrusters are installed at their individual positions according to their function, with the apogee engine installed under the lower platform together with the helium tank. Some of the 10 N thrusters for orbit control can be used as redundancy for the apogee engine. They are qualified for the very high firing durations required in such a failure situation.

Dual-Mode Propulsion System

Dual-mode propulsion systems **combine** the high performance of bipropellant apogee engines with the reliability and simplicity of monopropellant systems. Consequently, hydrazine is used as propellant instead of MMH. This requires the development of a new type of apogee engine.

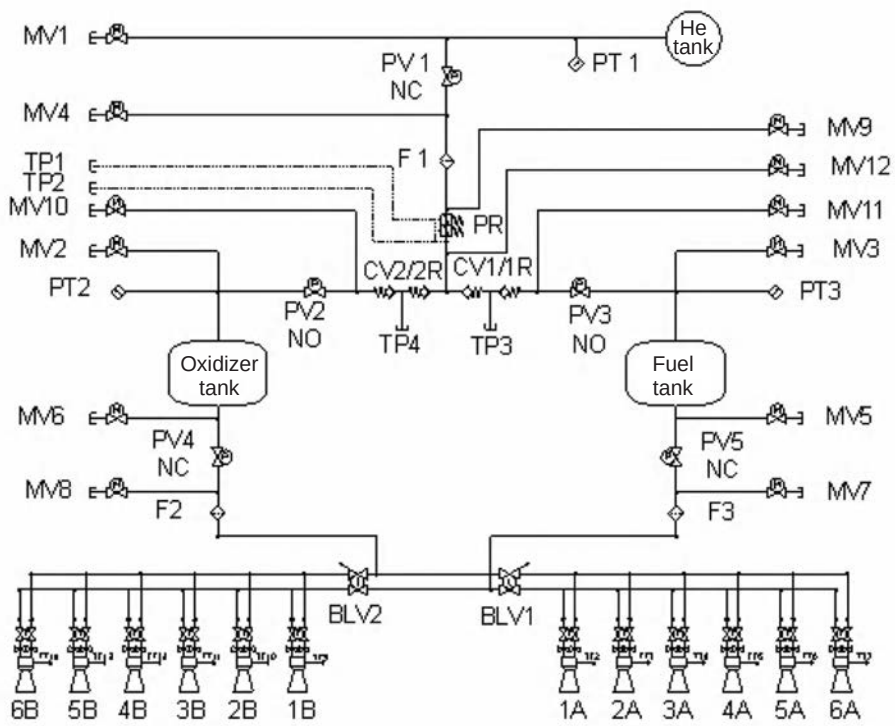


Figure 4.4.34: *Flow schematic of a bipropellant propulsion system (communications satellite) (Source: ASTRIUM).*

Attitude control is performed with hydrazine thrusters and if appropriate orbit control as well.

However, two more and even better alternatives for orbit control exist:

- The development of a 22 N bipropellant thruster also using hydrazine as fuel.
- The application of a hydrazine arcjet.

From a specific impulse point of view the latter is a much better solution but more expensive with respect to hardware cost.

The UK company Royal Ordnance (today AMPAC-ISP UK Ltd) has developed both types of thrusters, the apogee engine (e.g., LEROS 1b on the US Mars Odyssey and Mercury Messenger probes) and the 22 N orbit control thruster with N_2H_4 as fuel.

Dual-mode apogee engines have been developed in the USA and also in Japan. MON/N_2H_4 is a critical propellant combination, primarily because of the much lower temperature limit for thermal decomposition of the hydrazine fuel compared to MMH. This complicates drastically the development of such dual-mode engines (Figure 4.4.35). The specific impulse of this propellant combination is slightly higher than that of MON/MMH.

The ideal **mixture ratio** for MON/N_2H_4 thrusters is 1:1. However, the advantage of equally sized tanks is lost, as the demand for hydrazine is much higher than for MON.

Accordingly the tank volumes are either different or one oxidizer tank and two or more fuel tanks are used, as designed for NASA's Mars Global Surveyor, for example.

4.4.7 Electric Propulsion

Chemical propulsion systems are limited with respect to their specific energy content, as the reaction energy per mass unit is stored within the chemical propellants. This limits the achievable exhaust velocity and thereby the specific impulse. Electric propulsion systems are not limited with respect to their energy content, as the electrical power used to accelerate the propellant is supplied by an external energy source. Consequently the propellant can be accelerated to

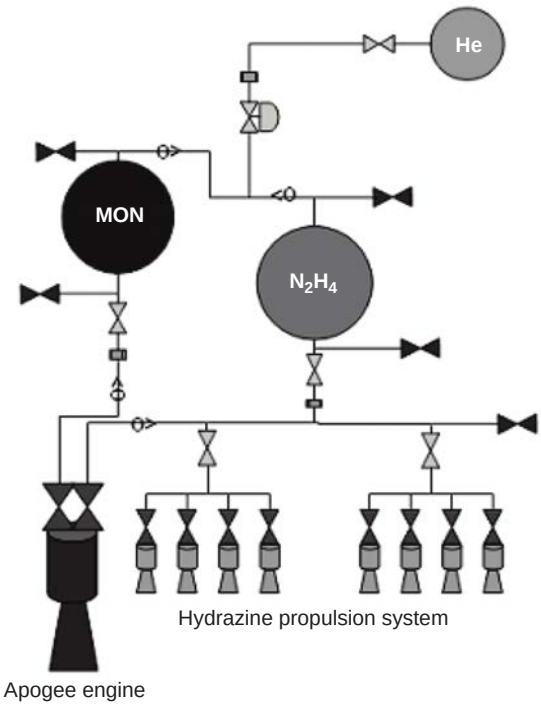


Figure 4.4.35: *Flow schematic of a dual-mode propulsion system.*

extreme velocities and thereby achieve very high specific impulses. However, the power as well as the thrust are limited by the available electrical energy delivered by batteries, solar generators or radioisotope thermoelectric generators (**RTGs**).

4.4.7.1 Thruster Types and Function

Three following main technology groups can be identified [4.4.26], [4.4.27], [4.4.28]:

- Electrothermal propulsion
- Electrostatic propulsion
- Electromagnetic propulsion.

Electrothermal thrusters were described in Section 4.4.6.1. The design and function of electrostatic and electrodynamic thrusters are described in the following.

For **electrostatic thrusters** the three main technologies using xenon as a propellant are described in the sketches of Figure 4.4.36.

In electrostatic thrusters the acceleration is caused by electrostatic potentials. Examples are the ion engines (types RIT, Kaufman or Hall effect), the colloidal thruster and the field emission thruster (type FEEP). The **ionization** can be performed by means of DC discharge, radiofrequency or electron synchrotron. The positively charged particles need to be neutralized by adding electrons outside of the acceleration zone.

In the RIT (**Radiofrequency Ion Engine**) (Figure 4.4.36, left), described for the first time by Horst Löb in 1962, the plasma is generated without the use of electrodes by means of a high radiofrequency field. This avoids the use of life limiting hollow cathodes within the discharge vessel. The acceleration of the ions is achieved by a grid system using a high-voltage potential in the range from 1200 to over 2000 V. Outside of the grid system, electrons are added by a neutralizer in order to neutralize the ion beam, thus avoiding the charging of the satellite structure. In total two xenon feed lines are required, one for the thruster and one for the neutralizer. Thrusters of the RIT type can achieve specific impulse values a factor 10 to 20 times higher than those of chemical propulsion systems. The specific impulse amounts to greater than 35 000 N s/kg at thrust levels between 0.1 and 250 mN. A RIT-10 operated at 15 mN after a launcher injection

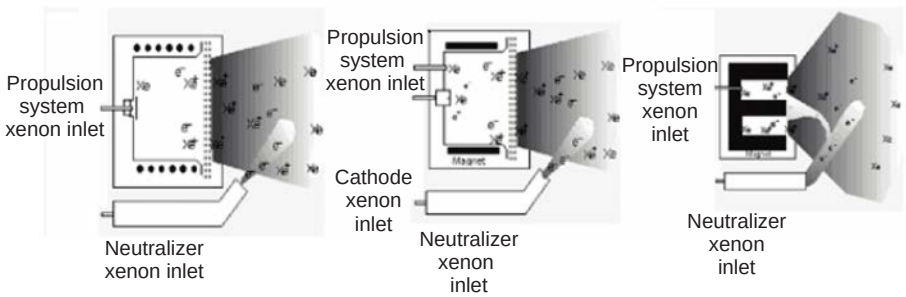


Figure 4.4.36: *Design of radio-frequency, Kaufman and Hall effect thrusters.*

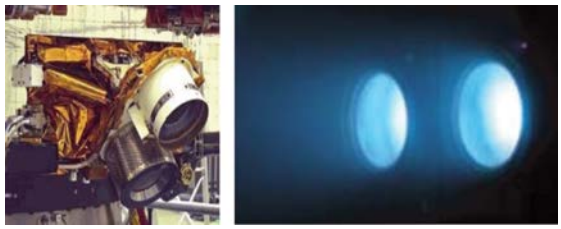


Figure 4.4.37 *RIT-10 and T5 (left) 2 Ion engines RIT 22 (right) of the ARTEMIS satellite (Source: ASTRIUM).*

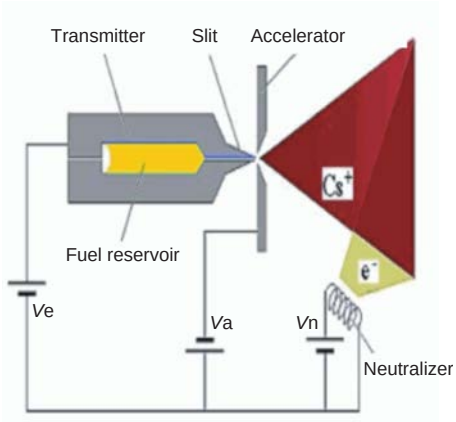


Figure 4.4.38 *FEEP thruster, slit design.*

failure succeeded in carrying ESA's Artemis satellite to its GEO position (Figure 4.4.37).

The **Kaufman thruster** is an electrostatic thruster as well; its development started as early as 1960. In a Kaufman thruster (Figure 4.4.36, middle) the plasma is generated by a plasma discharge between a hollow cathode and an anode. As in the RIT the ions are accelerated by a DC grid system. A magnetic field confines the plasma to increase ionization efficiency. In contrast to the RIT system, three xenon feed lines are required, creating more complexity for the thruster design and the overall system. Kaufman engines from the L-3 Corporation are used on the Boeing Space Systems satellite buses 601HP and 702. In addition, a 30 cm type was successfully used as primary propulsion for Deep Space 1 and is now being used on Dawn.

Both thrusters of RIT and Kaufman type are characterized by a very narrow ion exhaust plume angle of 25°.

The **field emission thruster** (FEEP) is another advanced electrostatic thruster using liquid metal ions to deliver very low thrust levels in the micronewton range at high specific impulses in the range of 60 000 N s/kg. In addition it allows thrust control with very high accuracy at very low thrust roughness. However, it requires high quality with respect to chemical purity and vacuum.

The low thrust levels make this technology a first choice for applications requiring an undisturbed microgravity environment or very accurate maneuvering between two satellites. The needle design, using indium as a propellant, consists of an emitter with an indium reservoir and a tungsten needle with a tip sharpness of a few micrometers.

Another FEEP technology uses cesium as a propellant that is transported by capillary forces between two very sharp, nearly parallel blades having a thickness of about 1 μm at the end (Figure 4.4.38).

For both FEEP types the emitters are supplied by a high positive potential of 3 to 12 kV. The high potential causes the liquid metal to form a so-called Taylor cone that has a very sharp tip with an extreme field strength of over 100 V/m. The acceleration of the ions is achieved with an extractor electrode that is at a potential of 1 kV. The resulting ion beam delivers a thrust between 0.1 and 300 μN [4.4.31].

Hall Effect Thrusters

In a **Hall effect thruster** (HET) or stationary plasma thruster (SPT) (Figure 4.4.36, right) xenon ions are accelerated by a DC field as well; however, contrary to the RIT or to the Kaufman thruster, not only acceleration grids but also the potential difference between the anode at the bottom of the circular discharge vessel and the external cathode in the range of a few hundred volts are used. Ionization is performed by a DC discharge between anode and cathode within the circular discharge vessel. By means of a magnetic field the electrons are trapped and accelerated in the discharge vessel to increase the ionization efficiency and form a Hall current. The heavy ions are not affected by this magnetic field.

HETs of different sizes can deliver thrust levels between 20 and 1000 mN. For example, the SPT100 manufactured by the Design Bureau Fakel in Kaliningrad with an electrical power consumption of

1500 W and a specific impulse of about 16 000 N s/kg produces a thrust of about 80 mN. In contrast to the gridded ion thruster, the ion beam of an HET has a relatively unfavorable divergence angle of 45° half angle [4.4.30].

Based on the ideas of A. Morozov, the development of the HET was started in the 1960s in the former USSR at TsNIIMASH and the Keldysh research center. Fakel was responsible for the industrialization. More than 200 HETs are operating in space.

ESA's Moon mission SMART-1 was propelled from GTO into lunar orbit by a PPS 1350 from Snecma Moteurs (Figure 4.4.39).

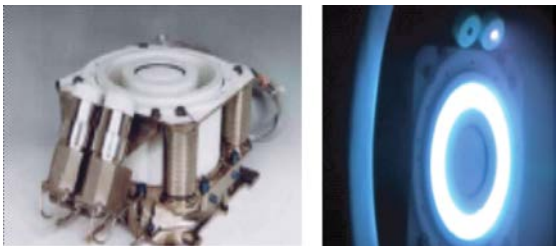


Figure 4.4.39: *Snecma PPS 1350 Hall effect thruster (Source: Snecma Moteurs).*

Electromagnetic Propulsion

In the low-thrust range, **pulsed plasma thrusters** (PPTs) provide extremely low impulse bits with a pulse frequency depending on the available power to charge the capacitor. The PPT system produces short energy pulses that vaporize, ionize and accelerate the solid propellant, normally Teflon (Figure 4.4.40).

Due to the pulsed operation with a frequency of 1 Hz or less, the average power demand is only a few watts. The energy is stored in a capacitor for about a second; after ignition the pulse length is only a few microseconds. The capacitor is then recharged during the times between pulse ignitions.

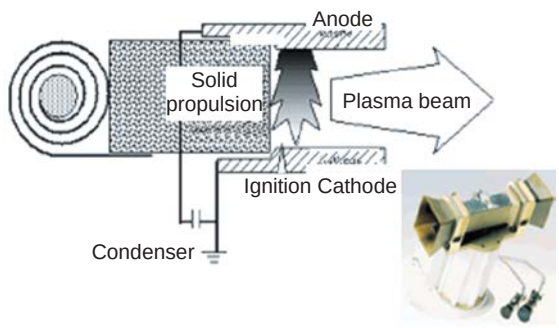


Figure 4.4.40: *Sketch and hardware of a pulsed plasma thruster (PPT) (Source: Aerojet).*

Magnetoplasmadynamic Thrusters

On the high-power side from tens of kilowatts to megawatts magnetoplasmadynamic (MPD) thrusters may be used for future human space exploration missions. In Germany MPD thrusters are under investigation at the Institute of Space Systems (IRS) at Stuttgart University.

The MPD thrusters use the Lorentz force (a force resulting from the interaction between a magnetic field and an electric current) to generate thrust.

In the 100 kW to 1 MW range the magnetic field that forms the magnetic nozzle to accelerate the ions is generated by the interaction of the DC discharge between the cathode and anode. This type is called the MPD self-field thruster.

In the lower power range from 10 to 30 kW the power is too weak to generate by itself a magnetic field for the generation of a magnetic nozzle. Consequently, a magnetic nozzle is generated by dedicated magnetic rings which create the applied MPD field. A great number of propellants such as xenon, neon, argon, hydrazine and lithium have been used, with lithium providing the best efficiency.

MPD thrusters with a laval nozzle are a combination of a thermal arcjet thruster and the MPD accelerator. At low currents most of the thrust is generated by the heating of the propellant in the DC discharge arc and the following almost ideal adiabatic expansion in the nozzle.

4.4.7.2 Electric Propulsion Systems

The following statements about electric propulsion properties can be made:

- The achievable **acceleration** is in the milli–g range.
- The **mass** of an electric propulsion system including the propellants can be a tenth of that of a chemical propulsion system, especially for high total impulse requirements.

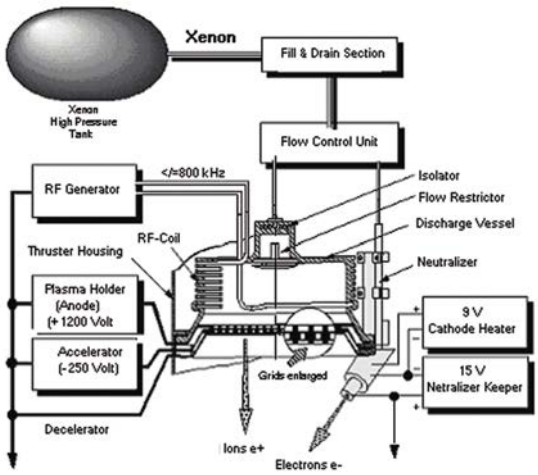


Figure 4.4.41: *Flow schematic of an RIT electric propulsion system (Source: ASTRIUM).*

- Due to the limited available power, the **transfer** from GTO to GEO performed with electric propulsion can take weeks to months. In most cases this delay does not fit into the operational concept of the satellite owner.

Almost all electric thrusters today are operated with xenon. It is stored under supercritical conditions as a liquid at a density of 1.83 g/cm^3. A typical storage pressure is 150 bar at 60 °C.

The block diagram of an electric propulsion system is visualized in Figure 4.4.41 for a RIT system. Besides the high-pressure tank and the thruster itself, two major subassemblies belong to the propulsion system.

The first is a mechanical or electromechanical **flow control unit**. As a completely mechanical device it consists of a pressure regulator (Figure 4.4.42) and a flow limiter. As an electromechanical device it consists of solenoid valves, pressure transducers, a small plenum tank and control electronics, all combined in an electromechanical flow regulation system.

The second is an electronic unit consisting of the high-frequency electronics, the high voltages to facilitate the ion acceleration, and the circuitry for power supply control of the engine itself.

A third subassembly, not shown in Figure 4.4.41, is the thruster orientation mechanism necessary in most applications for thruster realignment during the mission. The shift of the center of gravity of the spacecraft due to continuous propellant consumption can amount to several degrees. Therefore, engines are periodically realigned by means of the mentioned mechanism, as in the case of the Artemis satellite (Figure 4.4.37).



Figure 4.4.42: *Redundant pressure regulator (Source: VACCO Industries).*

A critical aspect of electric propulsion systems is the **electromagnetic compatibility** due to the high voltages and the required high power demand of the engines. Nevertheless, standard shielding and filtering methods are in general sufficient.

Due to their high performance and efficiency, electric propulsion systems are mostly applied for north–south station keeping of large communication satellites. Astrium's RIT 10 and the UK T5 are onboard Artemis. Boeing installs xenon ion propulsion systems (XIPS) from the L-3 Corporation on its communication satellites. The European satellite systems Spacebus 4000 of Thales-Alenia Space and Eurostar 3000 of Astrium use the SPT 100 HET.

In preparation of the very ambitious European LISA (Laser Interferometer Space Antenna) project, planned to measure gravitational waves as predicted by Einstein, a precursor satellite named LISA-Pathfinder will be equipped with an FEEP system.

4.4.8 Components for Chemical Propulsion Systems

Besides the most important propulsion components, the propellant tanks and the thrusters, the remaining

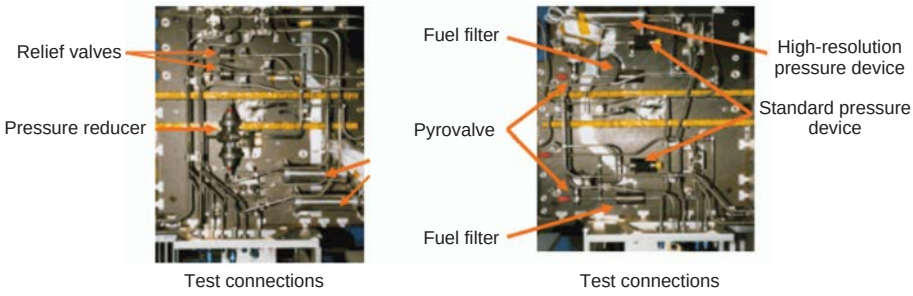


Figure 4.4.43: *Pressure control assembly (left) and propellant isolation assembly (right) (Source: ASTRIUM).*

components are briefly mentioned in this section for better understanding, but not dealt with in detail.

In propulsion systems built up in modules, these components are preassembled in so-called subassemblies depending upon their function in the system.

These subassemblies are (see Figure 4.4.43): (a) the **pressure control assembly** (PCA) for pressurization of the propellant tanks; and (b) the **propellant isolation assembly** (PIA) for isolation of the thruster branches.

In a geostationary communication satellite the PCA has completed all its tasks after the apogee injection maneuver. In an interplanetary probe such as Cassini, for example, it has to remain active up to the completion of the multiyear mission. Therefore the important pressure regulator is equipped with internal redundancy.

The high-accuracy **pressure transducers** installed in the PCA are used during the mission to determine the remaining propellant by means of the PVT method. The accuracy of these sensors amounts to 0.02% of the maximum pressure compared to 0.3% of the standard pressure transducers. The more accurate the determination of the propellant mass, the longer the satellite mission can be extended. The last few kilograms of propellant are used to raise the satellite from its valuable operating position in geostationary orbit into a so-called "graveyard" orbit.

The **pyrotechnic valves** exist, as described in Section 4.4.6.2, in a normally open (NO) and a normally closed (NC) version (Figure 4.4.44). The major advantage of this component compared to a solenoid valve is that it guarantees complete leak tightness in the closed condition.

Due to this feature, this type of valve is best suited to isolate certain areas of a propulsion system. This is



Figure 4.4.44: *Pyrotechnic valve (Source: ASTRIUM).*

done on the ground for safety reasons, while handling a satellite loaded with propellants or during the mission for isolation of sections no longer used, such



Figure 4.4.45: *Propellant and gas filters (Source: VACCO Industries).*

as an apogee engine after successful injection into geostationary orbit.

Filters are used in both the PCA (helium filter) and PIA (propellant filter). The type of filter presented in Figure 4.4.45 consists of chemically etched titanium disks assembled in packets of 500 or more under extreme clean-room conditions. The unit is completely sealed and guarantees a very high filtration quality. Usually filters with a filtration rate of 2 to 40 microns absolute are implemented for protection of components sensitive to contamination, such as thruster flow control valves.

4.4.9 Ground Support Equipment and Services

The challenges for the suppliers of propulsion systems are not limited to the development and production of components or complete propulsion systems as well as assembly, integration and test, but also include the design, development and production of the relevant ground support equipment for propellant loading and unloading, as well as propulsion system activities during launch preparation and immediately after launch.

Propellant Processing

While MMH can be used as delivered by the supplier, the oxidizer, delivered as N_2O_4, has to be mixed with NO to achieve MON 1 or MON 3.

Hydrazine, when delivered as "Monopropellant Grade" quality, has to be subjected to several distillation processes to get "High-Purity Grade" quality; in particular, the contents of aniline, water and iron have to be reduced.

Equipment for Testing and Satellite Servicing

Ground support equipment (GSE) of different sizes is required for tests on the component, subassembly and subsystem or system levels. Items belonging to this type of equipment are alignment tools for thrusters, used to keep the misalignment of the thrust vector to a minimum.

Equipment is needed for the **proof pressure and leakage tests** of the high- and low-pressure parts of the propulsion system, for leakage tests on component and subsystem levels during acceptance tests, and on the spacecraft level and during the launch campaign as well.

Each individual welding seam has to be inspected by **X-rays** and in most cases by penetrating inspection under ultraviolet light.

For satellite propellant and pressurant loading a considerable amount of GSE is needed. This type of GSE as presented in Figure 4.4.46 is used not only during relevant launch campaigns, but for loading and unloading propellant simulants (isopropyl alcohol or deionized water) during system-level vibration or acoustic noise tests as well.

Figure 4.4.46 shows the complete set of equipment needed for the propellant and pressurant loading of a bipropellant propulsion system consisting of propellant transport vessels; the loading carts for



Figure 4.4.46: *Bipropellant loading equipment (Source: ASTRIUM).*



Figure 4.4.47: *Unit tester for electric tests (Source: ASTRIUM).*

oxidizer, fuel and for pressurization with helium; as well as the transport containers for all the equipment. Similar equipment is needed for xenon or hydrazine loading.

The personnel engaged in the **propellant loading activities** need a complete set of **safety equipment** such as acidproof escape suits with autonomous breathing and communication devices, if not provided by launch site organizations.

The effort to **transport** propellants and the GSE to the different launch sites all over the world is considerable. Transportation by air is not feasible because of stringent safety regulations. Significant differences exist between the procedures and processes to get the GSE and propellants to launch sites such as Kourou, Kennedy Space Center or Baikonur. The relevant procedures have to be studied and respected to avoid surprises at the launch sites. The individual procedures and processes are not discussed further in this handbook.

A number of components such as pressure sensors, flow control valves and latching valves require electrical testing in turn with special test equipment. These tests include:

- Calibration of pressure transducers
- Measurement of opening and closing response times of valves
- Measurement of ohmic resistance of heaters, catalyst beds or valve heaters, etc.
- Verification of temperature sensor outputs.

In general the equipment needed for this type of testing is assembled into one single test device, known as the "unit tester" (Figure 4.4.47).

Bibliography

[4.4.1] Brown, C.D. *Spacecraft Propulsion*, AIAA Education Series. Reston, VA: AIAA, 1996 (incl. software).

[4.4.2] Erichsen, P. Performance Evaluation of Spacecraft. Propulsion Systems in Relation to Mission Impulse Requirements. Proceedings of the Second European Spacecraft Conference, *ESA SP-398*. Noordwijk: ESTEC.

[4.4.3] ASTRIUM ST. Spacecraft Propulsion, Hydrazine – Bipropellant Thrusters, Ion Propulsion. http://cs.space.eads.net/sp, 2008.

[4.4.4] Institut für Raumfahrtsysteme IRS. Grundlagen der Raumfahrtsysteme Vorlesung WS 2004/2005 und Vorlesung SS 2006 zur Lage- und Bahnregelung von Raumflugkörpern.

[4.4.5] Rocket Research Corporation. Monopropellant Hydrazine Design Data.

[4.4.6] TRW Systems. Propulsion System Thrusters and Components.

[4.4.7] Aerojet, Redmond Operations. Hydrazine Handbook.

[4.4.8] The University of Michigan. The Past and Future of Rocket Engine Propulsion Part II: Electrothermal Systems.

[4.4.9] Schmitz, H.D., Tiedtke, P. Qualification Test Results of German "KC IR 12 GA" Catalyst for Spontaneous Hydrazine Decomposition. *AIAA-1977-848*, 1977.

[4.4.10] Aerojet. MR 501 B Electrothermal Hydrazine Thruster. http://www.astronautix.com/engines/mr501b.htm, 2008.

[4.4.11] Wucherer, E., Cook, T., Stiefel, M. Hydrazine Catalyst Production – Sustaining S-405 Technology. *AIAA-2003-5079*, 2003.

[4.4.12] ARDE Inc. Composite Overwrapped Pressure Vessels – company specification. http://www.ardeinc.com. 2008.

[4.4.13] ARDE Inc. Propellant Tankage – company specification http://www.ardeinc.com. 2008.
[4.4.14] ARDE Inc. All Metal Pressure Vessels – company specification http://www.ardeinc.com, 2008.
[4.4.15] ASTRIUM ST. Spacecraft Propulsion, Propellant Tanks, Valves, Regulators. http://cs.space.eads.net/sp, 2008.
[4.4.16] ATK PSI Operations: Propellant Management Device Conceptual Design and Analysis. Vanes. *AIAA 91-2172*, 1991.
[4.4.17] ATK PSI Operations. Design & Development of a Communications Satellite Propellant Tank. *AIAA 95-2529*, 1995.
[4.4.18] ATK PSI Operations. http://www.psi-pci.com, 2008.
[4.4.19] Benedic, F., Leard, J.-P., Leloch, Ch. Helium High Pressure Tanks at EADS Space Transportation: New Technology with Thermoplastic Liner. 1st Symposium on Potentially Disruptive Technologies and their Impact in Space Programs, Marseilles, July 4–6, 2005.
[4.4.20] Vaughan, Ch. Propulsion Overview. Presentation, aa.washington.edu.
[4.4.21] Surrey Satellite Technology Ltd. Water Resistojet for Small Satellites. *SSTL-9003-03*, February 26, 2002.
[4.4.22] Surrey Satellite Technology Ltd. Nitrous Oxide Resistojet. *SSTL-9002-04* April 26, 2006.
[4.4.23] Anflo, K., Persson, S., Thormählen, P. *et al.* Flight demonstration of an ADN-Based Propulsion System. Proceedings of the 57th IAC/IAF/IAA (International Astronautical Congress), Valencia, Spain, October 2–6, 2006, *IAC-06-C4.1.08*.
[4.4.24] Smith, W.W., Smith, R.D., Davies, S.D. *et al.* Low Power Hydrazine Arcjet Flight Qualification. Presented at the 23rd International Electric Propulsion Conference, *IIWC-93-48*, 1993.
[4.4.25] ESA Science & Technology. Electric Spacecraft Propulsion. http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=34201, 2003.
[4.4.26] Martinez-Sanchez, M., Pollard, J.E. Spacecraft Electric Propulsion – An Overview. *J. Propul. Power*, **14** (5), 688–693, 1998.
[4.4.27] Pollard, J.E., Jackson, D.E., Marvin, D.C. *et al.* Electric Propulsion Flight Experience and Technology Readiness. *AIAA Paper 93-2221*, 1993.
[4.4.28] Wikipedia: Hall Effect Thrusters, Ion Thrusters and Spacecraft Propulsion. http://en.wikipedia.org/wiki/Hall_effect_thruster, 2008.
[4.4.29] ASTRIUM GmbH ST. Spacecraft Propulsion, Launch Support. http://cs.space.eads.net/sp/, 2008.
[4.4.30] Welander, B. Life and Operating Range Extension of the BPT-4000 Qualification Model Hall Thruster. *AIAA 2006-5263*, 2006.
[4.4.31] Genovese, A. 5000h Endurance Test of an Indium FEEP 2×2 Cluster. *AIAA 2006-4827*, 2006.
[4.4.32] Fischer, H.-M. *Europäische Nachrichten-Satelliten – Von Intelsat bis TV-Sat.* Lemwerder: Stedinger Verlag, 2006.

4.5 Attitude Control

Christian Arbinger and Bernard Lübke-Ossenbeck

The following section describes, after an explanatory overview of the field of **attitude determination** and control, the various mission requirements affecting the attitude control system. Different ways of describing the attitude, like the direction-cosine matrix, Euler description and quaternions, are presented. This is followed by a description of attitude dynamics, including an explanation of Euler differential equations, existing disturbance torques and attitude prediction. Different attitude sensors and actuators are also presented. The section concludes with practice-oriented methods of attitude control verification.

4.5.1 Introduction and Overview

The **attitude control system** (ACS) determines and controls the orientation of a vehicle in space. The attitude (or the orientation) is always described with respect to a reference system, either direction vectors within a reference coordinate system or as angles to defined reference axes.

The attitude of a spacecraft is generally independent of its position. Therefore, attitude control has to be clearly distinguished from orbit control.

As in every control aspect, **attitude control** consists of a **closed control loop** with:

- Measurement and determination of the actual attitude
- Comparison with the desired attitude
- Driving of actuators through a dedicated controller in order to achieve the desired attitude.

Within this control loop, the target could be either to achieve and keep a certain attitude, or to control the rate of an attitude change.

The execution of attitude control can occur in three different modes:

1. Under safety aspects as robust, coarse control (**safe mode**)
2. With the target of best performance as fine control (**science mode**, **precise mode**)

3. As a **transfer mode** in order to damp and control the spacecraft attitude rates or for attitude target acquisition.

4.5.1.1 Impact on the Satellite Bus

The Attitude Control System (ACS) is an important driver of system design. The mission goal (through the payload) provides the boundary conditions for the choice of appropriate **attitude sensors and actuators**. However, the basic functionalities of the satellite have to be ensured, in order to:

- Safeguard the performance of the **power system** (stable orientation of the solar arrays with respect to the Sun).
- Safeguard the **thermal conditions** (orientation control of the satellite body with respect to the Sun and the Earth).
- Adjust and maintain the attitude for maneuver (such as for orbit control).

In order to achieve a reliable ACS, the attitude requirements have to be considered in the satellite design process (e.g., with respect to mass distribution, potential oscillations of the structure, etc.).

4.5.1.2 Impact on the Payload

The functionality and performance of most payloads are closely linked to attitude determination and control: for example, for the orientation of a telescope for Earth observation, the orientation of the antennas of a telecommunication satellite, or the orientation of scientific payload.

For these tasks the attitude accuracy and stability play an important role and often the ability to carry out predefined **slew maneuvers**.

In many cases, telemetry data from the ACS is postprocessed together with the payload data in order to optimize the performance of the payload.

4.5.1.3 Application Areas

An attitude and orbit control system is implemented on nearly all space missions (whether human or unmanned). A distinction is made between "passive" and active attitude control. "Passive attitude control" means the steering of the attitude without the use of mechanical actuators, unlike active attitude control, where actuators are used within a control loop. The requirements, placed on the type of attitude control, depend on the mission characteristics.

During the mission, the attitude of the satellite is determined on-board in real time, for the various operational modes (safe mode, science mode, etc.). Because it is based on available on-board real-time data, this attitude control is not as accurate as precise attitude determination, which is performed during postprocessing on the ground. Additional data is used on the ground in order to minimize the sources of error which exist in the attitude determination process. The scientific payload of the satellite (e.g., telescopes, radar sensors or optical cameras) generally requires precise on-ground postprocessing of the attitude data in order to support the maximum resolution of the instruments.

4.5.2 Requirements for the Attitude Control System

The type and design of the ACS is mainly determined by the functional requirements (Table 4.5.1) of the satellite bus and the payload.

Stability requirements are normally derived from the payload and are indicated as the allowed angle deviation within a fixed time interval.

Besides these functional requirements, there are requirements for the performance of the various processes, whereby the most important are:

- **Accuracy,** distinguishing between real-time accuracy and *post facto* attitude accuracy. The accuracy is represented as a plus or minus interval or, more often, as σ values. (e.g., "0.1°, 3 σ" means that in 99.7% of all cases the residuals are below 0.1°).
- **Stability** represents the accuracy of keeping attitude rates.
- **Agility** describes the rate of time for changeovers between two dedicated attitudes.

In addition, the design of the ACS is driven by the following general requirements and boundary conditions:

- Cost.
- Design lifetime, for satellites typically from 5 to 15 years.

Table 4.5.1: *Functional requirements for an attitude control system.*

Requirement from	Functional requirement	Quality
Power subsystem	Orientation of solar arrays, static	Rough
Thermal subsystem	Orientation of radiators	Rough
Orbit control system	Orientation of orbit thrusters Acceptance of large-disturbance torques If necessary, guidance during long maneuvers	Accurate
Optical payloads	Orientation of telescopes Slew and rotation maneuvers, target tracing Provision of attitude data for postprocessing	Very accurate
Communication payloads	Orientation of "high-gain antennas" If necessary, execution of slew maneuvers	Accurate
Scientific payloads	Orientation of instruments Slew and rotation maneuvers, target tracing Provision of attitude data for postprocessing	Rough to very accurate
Safe mode	Orientation of solar arrays Attitude stabilization	Rough
Acquisition mode	Damping of high rotation rates Initial attitude determination	e.g., 10°/s e.g., within 100 s
Satellite operations	Autonomy FDIR (Failure Detection, Isolation and Recovery)	e.g., over 3 days e.g., switching between different spacecraft operation modes

- Reliability of the complete ACS, for example 95% over 10 years.
- No single points of failure, meaning that the failure of any *one* component shall be manageable.
- Compatibility with other satellite subsystems, for example, with respect to electromagnetic aspects or vibrations.
- Boundary conditions with respect to mass, power consumption and thermal budget.
- Orbital parameter of the satellite.
- Customer requirements, exclusion of certain components.

4.5.3 Parameterization of the Attitude

The attitude of a spacecraft is defined as the **angular deviation** between the axes of two coordinate systems. Definitions of typical **reference systems** used in the space field follow. This section shows different attitude representation methods and describes the parameterization of the spacecraft attitude.

4.5.3.1 Reference Systems

For the mathematical representation of the attitude of a body it is necessary to define two three-dimensional Cartesian coordinate systems. These systems are designated as the **reference system** and the **body system**. While the reference system is independent of the attitude motion of the considered body, the body-fixed coordinate system is anchored within the body, following the motion of that body.

The mathematical meaning of the term **attitude** is the angular deviation of these two coordinate systems. The difference vector of the origins of these coordinate systems (from the origin of the reference system to the origin of the body systems) is not relevant for attitude representation. The focus is the rotation of the two systems against each other. Figure 4.5.1 illustrates this description. The reference system is represented

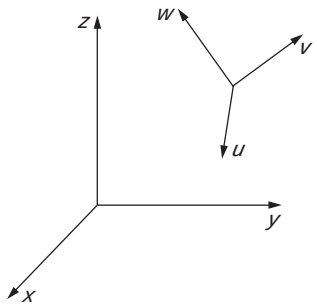


Figure 4.5.1: *Attitude representation with coordinate systems.*

by the vectors ***x***, ***y***, ***z***; the body system consists of the vectors ***u***, ***v***, ***w***.

When considering spacecraft in Earth orbits, the following reference systems are commonly used:

- **Inertial geocentric reference system:**
  - $+x$-axis points to the first point of Aries
  - $+z$-axis points to the North Pole of Earth
  - $+y$-axis forms a right-handed system with the $+x$- and $+z$-axes.
- **Earth-fixed geocentric reference:**
  - $+x$-axis points to the intersection of the equatorial plane with the Greenwich meridian
  - $+z$-axis points to the North Pole of Earth
  - $+y$-axis forms a right-handed system with the $+x$- and $+z$-axes.
- **Orbit reference system with its origin in the spacecraft's center of mass:**
  - $+z$-axis points to the geocenter (nadir direction)
  - $+y$-axis points in the direction to the negative orbit normal.
  - $+x$-axis forms a right-handed system with the $+y$- and $+z$-axes.

The body-fixed system, for symmetric spacecraft, is usually aligned with the **principal axes** of the satellite. Figure 4.5.2 shows a representation of such a system with the respective attitude angles. These angles will be defined in the following.

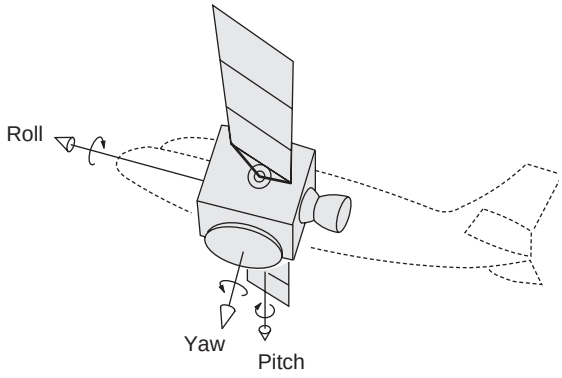


Figure 4.5.2: *Body-fixed coordinate system.*

4.5.3.2 Direction-Cosine Matrix

As defined in the previous section, the attitude of a spacecraft describes the orientation of a reference system with respect to a body system. The attitude is represented by a 3×3 transformation or attitude matrix ***A***.

The elements of this attitude matrix describe the vector dot product of the respective coordinate system axes and represent, therefore, the cosine of the angular deviation. The following equation shows this so-called direction-cosine matrix, based on the coordinate systems of Figure 4.5.1:

$$\boldsymbol{A} = \begin{bmatrix} u \cdot x & u \cdot y & u \cdot z \\ v \cdot x & v \cdot y & v \cdot z \\ w \cdot x & w \cdot y & w \cdot z \end{bmatrix} \tag{4.5.1}$$

This matrix allows for the transformation into the body system of vectors represented in the reference system. The following relation applies:

$$\boldsymbol{x}_B = \left[\boldsymbol{A} \right] \boldsymbol{x}_R \tag{4.5.2}$$

where:
x_R = vector in the reference system,
x_B = vector in the body system,
A = attitude matrix (orthogonal and symmetric).

Based on this representation, consecutive rotation sequences can be described by

$$A_{1-3} = A_{2-3}\, A_{1-2} \tag{4.5.3}$$

where:
A_{1-3} = complete rotation sequence from state 1 to 3,
A_{1-2} = rotation from state 1 to 2,
A_{2-3} = rotation from state 2 to 3.

With three consecutive rotations every coordinate system can be transformed into any other. The advantages of this attitude parameterization are that no singularities exist and no trigonometric/arithmetic operations are needed. As shown in Equation 4.5.3, there is a simple algorithm for rotation sequences.

However, this representation is of less importance in the field of space operations. During a space mission, the attitude system engineer in the control center checks the attitude data of the satellite on the telemetry screen after reception at the ground station. For this process, an attitude matrix representation is inappropriate, because the angular deviations between the two coordinate systems are not directly visible. In the field of satellite operations, it is common to use Euler angle representations, as described in the next section.

4

4.5.3.3 Euler Angles

For better visualization of the attitude, so-called Euler angles are used.

The three basic rotations around the respective coordinate system axes (x, y, z) can be described using **Euler angles,** as follows.

Rotation around the x-axis with angle φ:

$$\boldsymbol{A}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{bmatrix} \tag{4.5.4}$$

Rotation around the y-axis with angle θ:

$$\boldsymbol{A}_y = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \tag{4.5.5}$$

Rotation around the z-axis with angle ψ:

$$\boldsymbol{A}_z = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.5.6}$$

A special convention of these angles, derived from the steering angles in the field of aeronautics, is commonly used in the field of space engineering. The names of these angles are indicated in Figure 4.5.2. **Roll** defines an angle deviation around the longitudinal direction (x-axis). The **pitch** angle describes the rotation around the lateral axis (y-axis) and the **yaw** angle corresponds to a rotation around the vertical axis (z-axis). It is important to clearly define this angle assignment, in order to obtain an unambiguous description. In the design phase of a mission this should also be noted in the corresponding documentation to avoid misinterpretation.

4.5.3.4 Quaternions

Another method of attitude representation is to use quaternions. These hypercomplex numbers constitute an optimum with respect to memory load and numerical stability:

$$\boldsymbol{q} = \left[e_x \sin\left(\frac{\Phi}{2}\right) e_y \sin\left(\frac{\Phi}{2}\right) e_z \sin\left(\frac{\Phi}{2}\right) \cos\left(\frac{\Phi}{2}\right) \right]^T \tag{4.5.7}$$

where:
e_x, e_y, e_z = components of the unit vector along the rotation axis,
ϕ = rotation angle around this axis.

Analogous to the attitude matrix, rotation sequences can be easily expressed by

$$\boldsymbol{q}_{1-3} = \boldsymbol{q}_{1-2} \otimes \boldsymbol{q}_{2-3} \tag{4.5.8}$$

$$\boldsymbol{q}_{1-3} = \begin{bmatrix} q_{4(2-3)} & q_{3(2-3)} & -q_{2(2-3)} & q_{1(2-3)} \\ -q_{3(2-3)} & q_{4(2-3)} & q_{1(2-3)} & q_{2(2-3)} \\ q_{2(2-3)} & -q_{1(2-3)} & q_{4(2-3)} & q_{3(2-3)} \\ -q_{1(2-3)} & -q_{2(2-3)} & -q_{3(2-3)} & q_{4(2-3)} \end{bmatrix} \boldsymbol{q}_{1-2} \tag{4.5.9}$$

where:

$\boldsymbol{q}_{1-3}$ = complete rotation from state 1 to 3,
$\boldsymbol{q}_{1-2}$ = rotation from state 1 to 2,
$\boldsymbol{q}_{2-3}$ = rotation from state 2 to 3.

The disadvantage of quaternion parameterization is a poor physical illustration. Quaternions are a purely mathematical representation and they are not used for the visualization of attitude telemetry parameters during satellite operations. Nevertheless, in the on-board software and for attitude postprocessing, quaternions are commonly used.

4.5.4 Attitude Dynamics

4.5.4.1 Attitude Kinematics and Dynamics

The term **attitude** has now been defined and the different representation methods have been explained. For an attitude analysis it is important to understand the change of attitude by having a close look at **attitude kinematics** and **attitude dynamics**.

Kinematic Equations of Motion

The kinematic equation of motion is derived from observation of the attitude matrix change over time. This change is described as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[A\right]=\left[\Omega\right]\left[A\right] \tag{4.5.10}$$

with

$$\left[\Omega\right]=\begin{bmatrix} 0 & \omega_z & -\omega_y \\ -\omega_z & 0 & \omega_x \\ \omega_y & -\omega_x & 0 \end{bmatrix} \tag{4.5.11}$$

The elements ω_x, ω_y, ω_z represent the angular velocities around the corresponding axes of a body-fixed system.

For a kinematic consideration, attitude change is observed without the existing torques triggering this change. For kinematic attitude propagation, measurement and determination of the attitude rates are necessary. Dedicated sensors are described in Section 4.5.6.

Dynamic Equations of Motion

In contrast to the kinematic attitude description, dynamic attitude modeling also takes into account the acting forces and resulting torques.

Essential for dynamic attitude consideration is the definition of the relevant important physical units. The **angular momentum** represents the direction and velocity of a rotation around a reference point:

$$\boldsymbol{H}=\int_m \boldsymbol{r}\times\boldsymbol{v}\,\mathrm{d}m \tag{4.5.12}$$

where:
$\boldsymbol{H}$ = angular momentum vector,
$\boldsymbol{r}$ = position vector of a mass element,

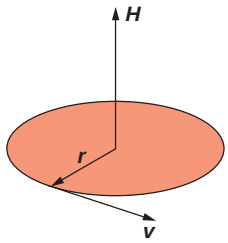


Figure 4.5.3: *Representation of angular momentum.*

$\boldsymbol{v}$ = velocity vector,
$\mathrm{d}m$ = mass element.

Figure 4.5.3 visualizes the angular momentum vector of a rotation.

Assuming that the satellite is a rigid body,

$$\boldsymbol{H}=\boldsymbol{I}\,\boldsymbol{\omega} \tag{4.5.13}$$

where:
$\boldsymbol{I}$ = inertial tensor,
$\boldsymbol{\omega}$ = angular velocity of the satellite.

A mathematically positive rotation of vector $\boldsymbol{\omega}$ can be determined with the right hand rule.

For the rotation of a body around its center of mass in an inertial coordinate frame, the following equation applies:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{H}=\sum\boldsymbol{M} \tag{4.5.14}$$

This equation means that the total change of angular momentum is equal to the sum of all torques $\boldsymbol{M}$ acting on the satellite body. For a body-fixed frame

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{H}+\boldsymbol{\omega}\times\boldsymbol{H}=\sum\boldsymbol{M} \tag{4.5.15}$$

Separated into components this equation becomes

$$\begin{aligned} \dot{H}_x+\omega_y\,H_z-\omega_z\,H_y&=M_x \\ \dot{H}_y+\omega_z\,H_x-\omega_x\,H_z&=M_y \\ \dot{H}_z+\omega_x\,H_y-\omega_y\,H_x&=M_z \end{aligned} \tag{4.5.16}$$

These equations represent the basis for dynamic attitude modeling, and also form the so-called Euler equations of motion.

The total torque $\boldsymbol{M}$ is the sum of torques acting on the spacecraft. The relevant disturbance torques are described in the following sections.

4.5.4.2 Disturbance Torques

In order to model the dynamic attitude of the spacecraft, the torques acting on the satellite have to be considered. A distinction is made between internal and external torques. **Internal torques** are usually generated by the actuators of the ACS (desired) and by movable mechanisms (generally not desired, e.g., fuel sloshing, structure mechanisms, solar panels). **External torques** are generated by the interaction of the space environment with the satellite. The corresponding magnitude of the disturbance torques is mainly related to the orbit and attitude of the satellite and the physical properties of the spacecraft. In the following only the most important disturbance torques are considered.

Gravity Gradient Torque

Starting from a satellite as an asymmetric rigid body with finite dimensions, an external torque acts on the spacecraft due to the force resulting from the Earth's gravity.

The gravity gradient torque can be represented in the body frame by the following relation:

$$\boldsymbol{M}_{\text{grav}} = 3\frac{GM}{r_S^3}\left[\boldsymbol{r}_S \times \left(\boldsymbol{I}_S\, \boldsymbol{r}_S\right)\right] \tag{4.5.17}$$

where:
GM = Earth's gravitational constant (see Section 2.2.1.1),
$\boldsymbol{r}_S$ = position vector of the satellite.

From this equation the following characteristics of the gravity gradient torque can be derived:

- The torque is perpendicular to the local vertical.
- The torque is inversely proportional to the cube of the geocentric distance of the spacecraft.
- The torque vanishes for spherically symmetric spacecraft with equal principle moments of inertia.

Solar Radiation Torque

The photons of solar radiation impacting the surface of a satellite create a force that results in a torque around the center of mass of the spacecraft. The **solar radiation pressure** is virtually independent of the orbital altitude of the satellite in an Earth orbit. The main factors driving this torque are:

- The intensity and spectral distribution of the incident radiation
- The surface geometry and the optical characteristics of the spacecraft
- The orientation of the Sun vector relative to the spacecraft.

The acting force corresponds to the impulse change of the incident light. For the simplified case of direct reflection it can be assessed by:

$$\boldsymbol{F}_{\text{sol}} = \frac{S}{c}\left(1+r\right) A \left(-\boldsymbol{e}_{\text{sun}}\right) \tag{4.5.18}$$

where:
S = solar constant (1371 W/m^2),
c = speed of light in vacuum,
A = projection of the satellite surface in Sun direction,
r = reflection coefficient,
$\boldsymbol{e}_{\text{sun}}$ = unit vector in Sun direction.

Based on Equation 4.5.18 the resulting solar radiation torque is

$$\boldsymbol{M}_{\text{sol}} = \left(\boldsymbol{r}_A - \boldsymbol{r}_S\right) \times \boldsymbol{F}_{\text{sol}} \tag{4.5.19}$$

where:
$\boldsymbol{M}_{\text{sol}}$ = solar radiation torque,
$\boldsymbol{F}_{\text{sol}}$ = force resulting from solar radiation pressure,
$\boldsymbol{r}_A$ = point of attack of force in center of area,
$\boldsymbol{r}_S$ = position vector of the satellite center of mass.

Aerodynamic Torque

For spacecraft in low Earth orbit (LEO), the interaction of the satellite's surface with the **residual atmosphere** cannot be ignored. The aerodynamic torque is the dominant environmental torque for spacecraft below 400 km altitude. The force due to the impact of atmospheric molecules on the spacecraft surface can be modeled as an impact without reflection.

The resulting disturbance torque is described as

$$\boldsymbol{M}_{\text{aero}} = \left(\boldsymbol{r}_A - \boldsymbol{r}_S\right) \times \boldsymbol{F}_{\text{aero}} \tag{4.5.20}$$

whereby

$$\boldsymbol{F}_{\text{aero}} = -\frac{1}{2}\,\rho\left(\boldsymbol{r},t\right) C_{\text{D}}\; A\,\dot{\boldsymbol{r}}^{2}\cdot\frac{\dot{\boldsymbol{r}}}{\dot{r}} \tag{4.5.21}$$

where:
C_D = aerodynamic coefficient,
A = satellite surface in flight direction,
$\rho(r,t)$= atmospheric density, a function of position and time,
$\dot{\boldsymbol{r}}$ = velocity vector of the satellite,
$\dot{r}$ = absolute value of the velocity vector.

This force always acts counter to the flight direction of the satellite.

Magnetic Disturbance Torque
This disturbance torque results from the interaction of the residual magnetic field (**dipole moment**) of the spacecraft with the Earth's magnetic field. During the satellite design phase, this magnetic residual field is minimized by appropriate shielding and by taking it into account when establishing the cabling layout of the satellite. The attitude control algorithm needs the absolute value of this residual moment also as input for dimensioning the actuating variable and in order to ensure the stability of the control loop. The resulting magnetic disturbance torque is calculated as

$$\boldsymbol{M}_{\text{mag}} = \boldsymbol{m}_{\text{S}} \times \boldsymbol{B}_{\text{Erde}} \tag{4.5.22}$$

where:
$\boldsymbol{m}_S$ = magnetic dipole moment,
$\boldsymbol{B}_{\text{Earth}}$ = Earth's magnetic field vector.
The rough ratio of the disturbance torques acting on a satellite at 1000 km is
$\boldsymbol{M}_{\text{grav}} : \boldsymbol{M}_{\text{mag}} : \boldsymbol{M}_{\text{sol}} : \boldsymbol{M}_{\text{aero}} = 1000 : 250 : 2 : 0.5$.

Other Disturbance Torques
The dominant external disturbance torques were described in the previous sections, but there are also internal torques which should be considered in the design of the attitude control. Internal disturbance torques are, for example:

- Leakage within the propulsion system
- Fuel sloshing in the propulsion tanks
- Moving parts of the satellite structure
- Crew movement on human space flight missions.

The modeling of these disturbance torques is very difficult. Nevertheless, the maximum values of these internal torques estimated during simulations should be taken into account when designing the attitude control loop in order to achieve stable and safe attitude control.

4.5.5 Attitude Determination and Control

4.5.5.1 Attitude Determination

Attitude determination is the process of calculating the actual **satellite attitude** (**three-axis attitude**) using appropriate measurements (see the section on attitude sensors). The raw measurements generated by the different sensors are combined after appropriate filtering. Based on this data the attitude determination algorithm is used to calculate the actual three-axis attitude of the spacecraft. Different attitude determination methods are used and are described in the literature. In this section algebraic attitude determination is presented as an example.

The algebraic method is based on the direct-cosine matrix representation of the attitude. Two vectors, $\boldsymbol{u}$ und $\boldsymbol{v}$, define an orthogonal coordinate system with the basis vectors $\boldsymbol{q}$, $\boldsymbol{r}$ and $\boldsymbol{s}$ according to

$$\begin{aligned}\boldsymbol{q} &= \boldsymbol{u}\\ \boldsymbol{r} &= \frac{\boldsymbol{u}\times\boldsymbol{v}}{\left|\boldsymbol{u}\times\boldsymbol{v}\right|}\\ \boldsymbol{s} &= \boldsymbol{q}\times\boldsymbol{r}\end{aligned} \tag{4.5.23}$$

assuming that $\boldsymbol{u}$ and $\boldsymbol{v}$ are not parallel.

For a given time, two vectors $\boldsymbol{u}_b$ and $\boldsymbol{v}_b$, measured by attitude sensors, form the body-fixed coordinate frame as described above and therefore the spacecraft attitude matrix $\boldsymbol{A}_b$:

$$\boldsymbol{A}_{\text{b}} = \left[\,\boldsymbol{q}_{\text{b}} \vdots \boldsymbol{r}_{\text{b}} \vdots \boldsymbol{s}_{\text{b}}\,\right] \tag{4.5.24}$$

These two vectors could be, for example, the Sun vector as measured by a Sun sensor, and the Earth vector as measured by an Earth sensor. The two vectors are needed in the defined reference system in relation to which the attitude is described. (To repeat: attitude is the angular deviation of a reference frame against a body-fixed frame.) These reference vectors, $\boldsymbol{u}_r$ and $\boldsymbol{v}_r$, form the reference matrix $\boldsymbol{A}_r$ with

$$\boldsymbol{A}_r = \left[\boldsymbol{q}_r \vdots \boldsymbol{r}_r \vdots \boldsymbol{s}_r \right] \tag{4.5.25}$$

As defined in Section 4.5.3.2, the direct-cosine matrix is determined by the following coordinate transformation:

$$\boldsymbol{A}\ \boldsymbol{M}_\mathrm{r} = \boldsymbol{M}_\mathrm{b} \tag{4.5.26}$$

This equation transforms the column vectors of $\boldsymbol{M}_\mathrm{r}$ into the column vectors of $\boldsymbol{M}_\mathrm{b}$. The equation can be solved for the direct-cosine matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{M}_\mathrm{b}\ \boldsymbol{M}_\mathrm{r}^{-1} \tag{4.5.27}$$

As $\boldsymbol{M}_\mathrm{r}$ is orthogonal, it follows that

$$\boldsymbol{A} = \boldsymbol{M}_\mathrm{b}\ \boldsymbol{M}_\mathrm{r}^{\mathrm{T}} \tag{4.5.28}$$

The algebraic method with two vectors is also referred to as **deterministic attitude determination**. The disadvantage of this method is that the attitude determination is incorrect if one of the measured vectors is wrong. But several measurements are generally available at the same time. By using the **least squares estimation method**, better and more stable attitude solutions can be achieved. For this estimation method a cost function is established and with the different measurements the minimum of this cost function is calculated and an attitude solution is generated. More complex methods make use of a Kalman filter for the attitude determination algorithm. Within this filter information from previous observations (measurements) as well as the dynamic characteristics are also taken into account in order to obtain an optimal estimation of the attitude parameter. Kalman filters for attitude applications can be found in the literature listed in the Bibliography.

For attitude determination, consideration of the **attitude prediction** (or **attitude propagation**) process is also very important. Within this process the attitude state is predicted for a future point in time by using the kinematic and dynamic equations of motions. This attitude prediction is important for attitude control in order to adapt the control variables for the upcoming control steps. Also within the ground satellite operations, the satellite attitude is propagated to allow for the generation of telecommands, for instance for orbit maneuvers or for scientific operations.

4.5.5.2 Attitude Control

The active modification and control of the satellite's attitude is understood as **attitude control**.

Figure 4.5.4 shows a typical design of an attitude **control loop.** The satellite should keep a desired attitude. This desired attitude can be either defined through an on-board algorithm or provided to the

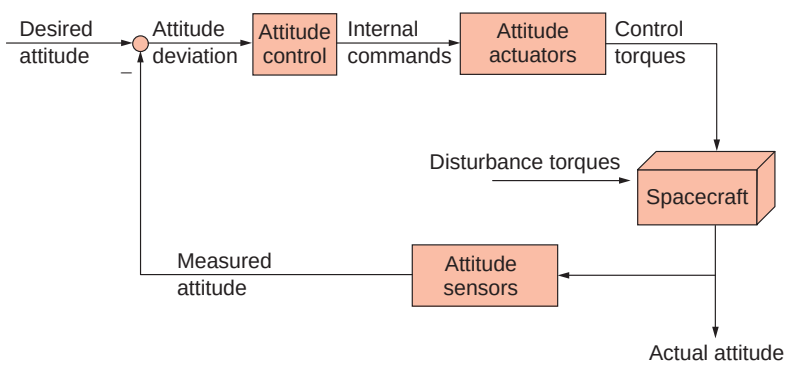


Figure 4.5.4: *Attitude control loop.*

satellite by telecommanding from the satellite operations team on the ground. The desired attitude is compared to the measured actual attitude of the spacecraft. The delta of this process is the **attitude deviation**. This deviation value is transferred to the attitude controller. Within this control algorithm the commands for the actuators are generated and transferred. The attitude actuators create a torque and actively change the attitude. In parallel the external and internal torques act on the satellite. The resulting attitude is measured by the sensors and the control loop is closed.

Different approaches exist for the attitude control algorithm. The most common approach is to use the so-called proportional, integral and derivative controller (PID controller). The PID controller reacts to a delta-signal ε (attitude deviation) as follows:

$$u(t) = K_P\, \varepsilon(t) + K_I \int \varepsilon(t)\, dt + K_D\, d\varepsilon(t)/dt \quad (4.5.29)$$

where:
$u(t)$ = controller output in the time domain,
$\varepsilon(t)$ = delta signal (attitude deviation),
K_P, K_I, K_D = PID controller parameters.

In the design phase of the spacecraft it is especially important to pay attention to the dimensioning and functionality of the ACS, because the success of a mission depends on the quality of the attitude control. For this reason the ACS is modeled with software and up-to-date satellite characteristics information is used. Section 4.5.8 takes a closer look at the verification of the ACS.

4.5.6 Attitude Sensors

4.5.6.1 General Aspects of Attitude Sensors

Attitude sensors provide measurements of the actual **attitude status**. Here the attitude is measured either as absolute (with respect to a reference coordinate system) or as relative information (e.g., dedicated angles or attitude changes).

Determination of the **absolute attitude** is based on the direction of two linearly independent vectors of a corresponding reference frame, as described above. The following vectors can be measured on-board:

- Earth magnetic field vector
- Direction vector to the Sun
- Direction vectors to stars
- Direction vector to the Earth (or the angle to the Earth's horizon)
- Direction vectors to the satellites of a global satellite navigation system (GNSS) such as GPS.

The accuracy of three-axis attitude determination strongly depends on the accuracy of the single measurements and the relative orientation of the vectors (optimal is 90°).

Besides direct absolute attitude determination, **angular differences** and **rotation rates** can be inertially determined.

The advantage of **inertial measurements** is based on the fact that the measurement values are available continuously and independently of external sources. Attitude determination with inertial sensors needs periodic alignment with absolute attitude measurements (Table 4.5.2).

Table 4.5.2: *Measurement methods of attitude sensors.*

Method	Example	Characteristics
Direct	Star sensor	Three axes, high accuracy
	GNSS attitude measurements	Three axes, medium accuracy
Indirect	Magnetometer Earth sensor Sun sensor	Simple method High reliability Transformation of measurements needed
Inertial	Gyroscope, Gyros	Attitude alignment High accuracy over short periods Very high angular resolution Measurement independent of external sources

Attitude Measurement Errors

As with every measurement, the result of attitude measurements contains systematic and random errors.

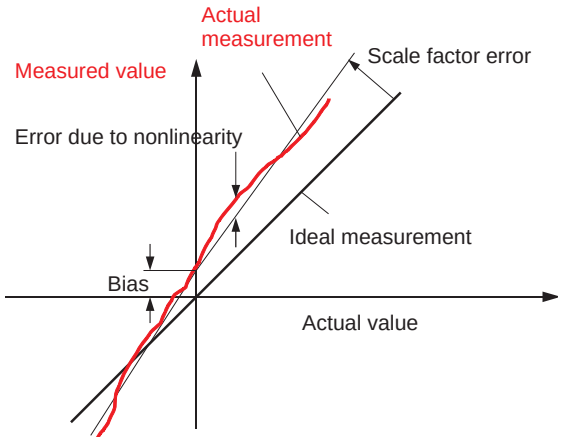


Figure 4.5.5: *Error parameterization.*

Systematic errors are usually consequences resulting from secondary effects (e.g., temperature, temperature gradient). If these consequences are known, systematic errors can be compensated to a large extent.

Random errors result from unknown or unpredictable effects on the measurements (e.g., measurement noise). Therefore they cannot be compensated, but can be reduced by averaging (or filtering). However, this is always a compromise between timeliness and accuracy.

Parameterization of Errors

Measurement errors are deviations from ideal sensor performance. They can be classified by the following parameters (Figure 4.5.5):

- Bias
- Scale factor error
- Nonlinearity
- Asymmetry (different scale factors for the positive and negative regime)
- Noise
- Quantification error.

Random errors result in general from the sum of different influences and are commonly described as the standard deviation (with the parameter σ).

Time Response

Another error source results from the time correlation of the measurement and plays a significant role for dynamic processes. Important variables are:

- **Dead Time:** Measurement value and time of measurement are shifted in time; this is in general an effect of digital signal processing.
- **Delay Time:** The measurement value is continuously determined during a certain time interval, during which this value is changing. Within this process changes are displayed with a delay and periodic characteristics are shown with smaller amplitude. The delay time is usually referred to as the "bandwidth" of the measurement. The bandwidth complies with the frequency at which half of the effective value (around 70% of the amplitude) is displayed.

4.5.6.2 Star Sensors

Star sensors, or star trackers, determine the inertial attitude through identification of **star patterns**. They consist of a camera (usually a separate sensor head) and a connected electronic part for image processing and attitude determination.

In order to determine the three-axis attitude, the identification of two stars is sufficient, whereas for initial attitude acquisition a pattern consisting of four stars is preferred to achieve unambiguous results. By comparison of the star pattern within the on-board star catalog, the attitude of the star sensor can be calculated.

Accuracy

The achievable accuracy is determined by:

- **Focal length** of the objective (typically 30–50 mm)
- **Pixel resolution** of the detector (typically 0.02°)
- **Signal strength** of the visible stars (visual magnitude mv 5 to mv 6.5, achievable with large apertures) and the noise of the detector
- The **focusing** and the interpolation method
- **Number** of processed stars (typically 3–15).

The geometry of the observation gives the result of a five to ten times higher accuracy around the axes perpendicular to the optical axis than around the optical axis. The accuracy of good star sensors is in the region of a few arc seconds.

Technical Design

In front of the objective there is a so-called **baffle**, a light trap which blocks the lateral incident stray light from the Earth and Sun (Figure 4.5.6). With baffle

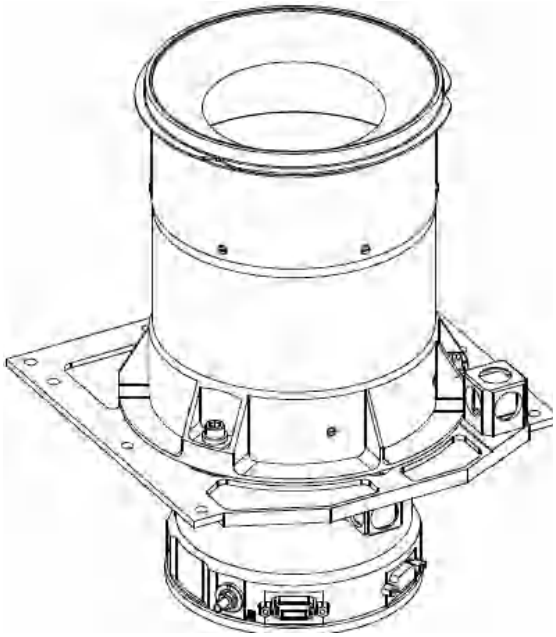


Figure 4.5.6: *Sensor head with baffle (Source: Jena-Optronik).*

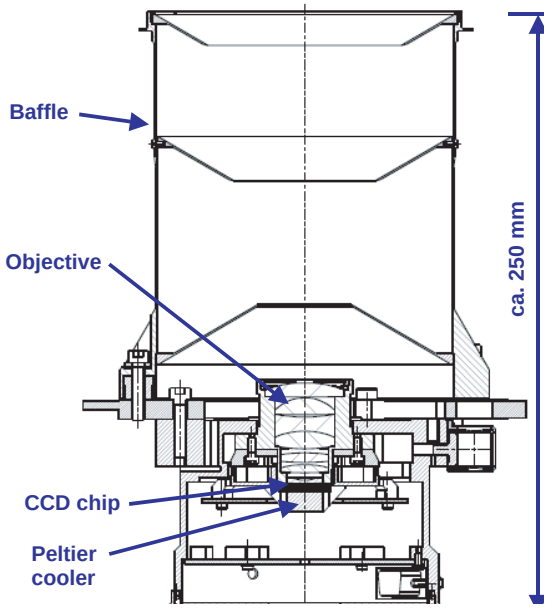


Figure 4.5.7: *Cross-section of a star sensor head (Source: Jena-Optronik).*

lengths of around 200 mm, the optical axis can be directed until about 30° to the Sun (**Sun exclusion angle**). Even more important is the **albedo exclusion angle**, representing the tolerance angle for reflective light from the Earth. For the above-mentioned baffle length, this angle is about 25°. This enables star sensors to operate in LEO within the antinadir hemisphere. Without a baffle, the weak starlight would, over wide angle ranges, be overlaid by stray light.

The objective projects the star light onto a **CCD chip** (Figure 4.5.7). The focal length is adjusted so that the light of a star is spread over several pixels. This makes it possible to determine the center point of the light distribution, which means the true position of the star on the chip, with subpixel accuracy.

CCD chips are on the one hand very light sensitive; on the other hand, they suffer from **hard radiation**, especially from the proton radiation of the Sun. This radiation can permanently damage the crystalline structures of the chip, increasing the noise floor on the CCD. This effect appears less at lower temperatures, therefore Peltier elements are frequently used for cooling.

Active pixel sensors (APSs) are more radiation resistant than CCD sensors, and also have a signal processing advantage by allowing to directly read out at dedicated positions.

Application Areas

Star sensors are the dominating technology for satellite attitude determination because of their accuracy and the three-axis attitude information. Installation of several sensors (or sensor heads), operating in different view directions, allows minimization of glare effects from the Earth, Sun and Moon. In addition, the accuracy of attitude determination can be further enhanced by careful selection and combination of the appropriate sensors and their respective axes (Table 4.5.3).

4.5.6.3 Sun Sensors

Sun sensors measure the direction to the Sun and therefore provide two-axis attitude information. They can be divided into coarse Sun sensors and more accurate fine Sun sensors.

Table 4.5.3: *Reference values for star sensors of medium accuracy.*

Attributes	Value	Annotations
Accuracy (EOL)	LOS 0.025°, 3σ Cross 0.003°, 3σ	EOL (End of Life) approx. 10 years, rotation rates < 0.1°/s LOS (Line of Sight) = optical axis
Temperature stability	±0.003°	Temperature range –30 to +30 °C
Update rate	8 Hz	–
Baffle	30° (Sun exclusion angle) 25° (albedo exclusion angle)	Angular distance between incident radiation and optical axis
FOV	About 18° × 13°	FOV = Field of View
Temperature range	–45 to +35 °C (sensor head) –45 to +50 °C (electronic box)	Temperature at the mounting interface
Mass	3.7 kg	Total: head, electronic box, harness
Power	8 W (cooler off) 14 W (cooler on)	At 20 °C ambient temperature

4

Coarse Sun sensors in general consist of solar cells which are attached to the different sides of the spacecraft to maximize the field of view (Figure 4.5.8). On the basis of measurements and comparison of the currents, generated by the different cells, the direction of highest illumination can be determined, whereby the disturbance caused by the Earth's albedo can be tolerated. With this simple measurement, attitude accuracies of 10 to 20° can be achieved, which are adequate for safeguarding the elementary thermal and power conditions for a satellite.

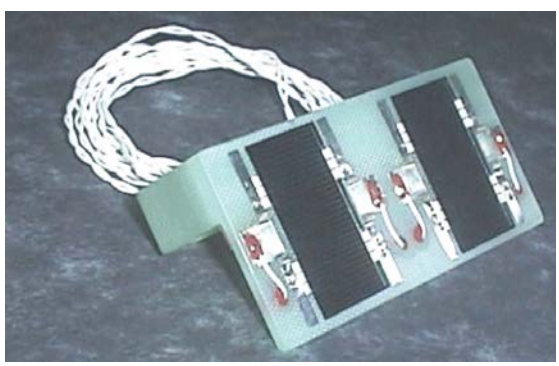


Figure 4.5.8: *Solar cells on a mounting structure as part of a coarse Sun sensor (Source: OHB-System).*

Simplicity, robustness and independence from the actual satellite attitude predestine this measurement principle for the attitude acquisition and safe modes.

Fine Sun sensors aim for an accurate measurement of the Sun's direction. The influence of diffuse albedo light can be removed to a large extent by directing the sunlight onto the detector through baffles and slits. The detector can be an array of photocells or a CCD chip. Accordingly, the field of view of these sensors is limited.

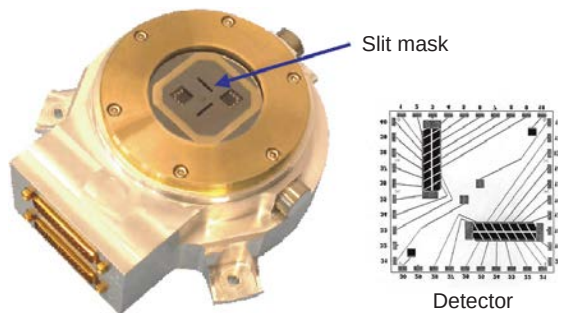


Figure 4.5.9: *Fine Sun sensor (Source: Jena-Optronik).*

Sun sensors are characterized (Table 4.5.4) by high reliability and provide (outside the Earth's shadow), even at high rotation rates, a stable attitude signal.

4.5.6.4 Earth Sensors

For most satellites, the payload is oriented toward Earth, whereas the satellite is oriented with respect to an orbit fixed frame. Earth sensors detect the Earth's horizon and provide the possibility to directly determine the attitude angles around the roll and pitch axes.

Table 4.5.4: *Typical values for Sun sensors.*

Aspect	Fine Sun sensor	Coarse Sun sensor
Attitude accuracy	0.01°	15°
Power consumption	1 W	0 W, connection to AD converter
Mass	1 kg	0.02 kg
Application area	Sun pointing	Orientation, safe mode

The measurement is performed in the thermal infrared spectrum at about 15 µm (CO_2 radiation band) and is therefore available over the whole orbit. In the case of the Sun or Moon crossing through the field of view, temporary disturbances are possible.

Besides the actual measurement accuracy, which is strongly dependent on temperature, the following systematic errors occur:

- The **irregular radiation distribution** within the Earth's atmosphere results in deviations of 0.1°, depending on the season.
- Through the **Earth's oblateness** the angle between the Earth's horizon and the Earth's center varies. Without compensation a deviation of up to 0.3° would result from this effect.

Static and scanning configurations of Earth sensors are used.

Static Earth Sensors

Static Earth sensors operate within the Earth's horizon staying **permanently** in the sensor's field of view. The Earth's thermal radiation is mapped with a germanium lens on two detectors (fields A and B of Figure 4.5.10). By comparing the generated voltages of these two fields, the position of the Earth's horizon within the sensor can be calculated.

The space field provides a reference value and shows in parallel the polarity of the measurement.

For **two-axis measurements** at least two sensor heads (Figure 4.5.11) have to gauge different sections of the Earth's horizon. Because these sensors have a limited field of view and are rigidly mounted on the satellite, this static measurement principle is less suitable in high elliptical orbits.

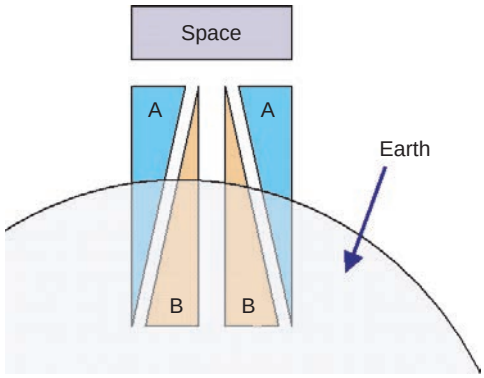


Figure 4.5.10: *Measurement principle of static Earth sensors.*



Figure 4.5.11: *Three static Earth sensors attached to a bracket. Advantages of this configuration are the simple thermal control, similar thermal conditions, two out of three redundancy, and compensation of offset errors (Source: Goodrich).*

Scanning Earth Sensors

A scanning Earth sensor passes the Earth's horizon through the detector. The scan is performed either by a **rotating mirror** (conical Earth sensor) or happens automatically for a spinning satellite.

In contrast to the static sensor, the changing transition of the horizon signal is measured; therefore the constant error components are eliminated within the signal. This results in higher accuracies for scanning Earth sensors than for static ones.

The time span between horizon signals corresponds to the angle around the **roll axis** of the satellite. The phasing of the horizon signals determines the angle around the **pitch axis** (see Figures 4.5.12 and 4.5.13).

Dedicated configurations exist for **three-axes-stabilized** satellites in GEO. Here the infrared radiation is mapped by an oscillating mirror on the detector.

Application Areas

The advantage of Earth sensors (see, e.g., Figure 4.5.14) is the direct measurement of the roll and pitch angles. If the satellite is equipped with a reaction wheel stabilizing the direction of the pitch axis, the angle around the yaw axis can be determined indirectly because the inertial attitudes of the roll and yaw axes are interchanged after a quarter orbit. The formerly unmeasurable error around the yaw axis is conserved by the wheel and then becomes measurable through the roll angle. This process becomes applicable, for example, for geostationary satellites.

The disadvantage of the Earth sensor is its limited accuracy. These sensors are being replaced more and more by star sensors, which provide a three-axis attitude signal with much higher accuracy (Table 4.5.5).

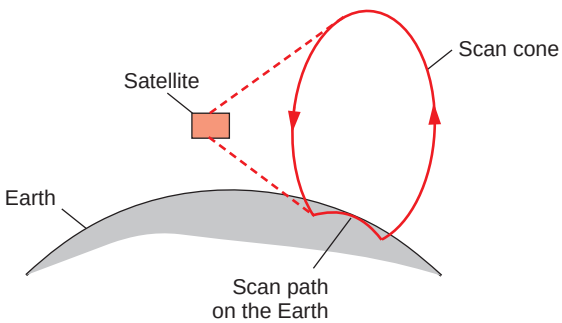


Figure 4.5.12: *Scan geometry (Source: Ithaco).*

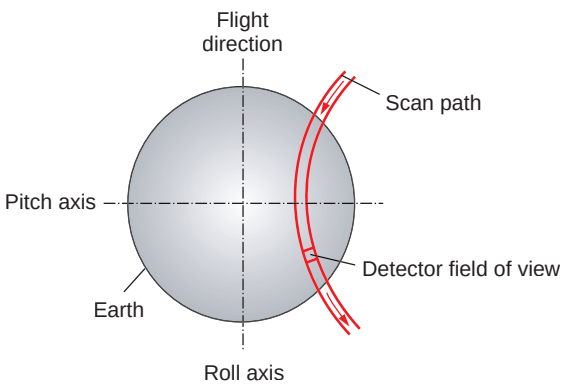


Figure 4.5.13: *Measurement principle of scanning Earth sensors; dependence of the detector signal on the roll and pitch axes of the satellite (Source: Ithaco).*

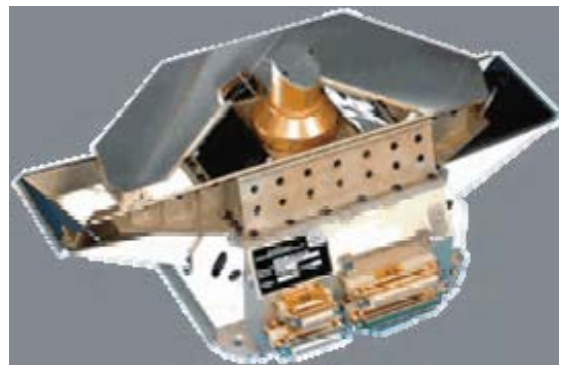


Figure 4.5.14: *Conical Earth sensor (Source: Sodern).*

Table 4.5.5: *Typical values of static and scanning Earth sensors.*

Aspects	Static	Scanning	Comments
Accuracy	1°, 3σ	Bias: 0.05°, 3σ Noise: 0.05°	After compensation for systematic errors
Field of view	20° (× 15°)	Roll ±30° Pitch ±20°	LEO
Measurement axes	1	2	–
Mass	0.2 kg	3.5 kg	–
Power consumption	0.35 W	7.5 W	Without thermal control

4.5.6.5 Magnetometers

The measurement of the Earth's **magnetic field** can be achieved easily by making use of several physical effects. Thus, magnetometers are in general cheap and reliable. For the most commonly used measurement principle, the fluxgate sensor (Figure 4.5.15), direction accuracies between 0.1° and 1° can be achieved

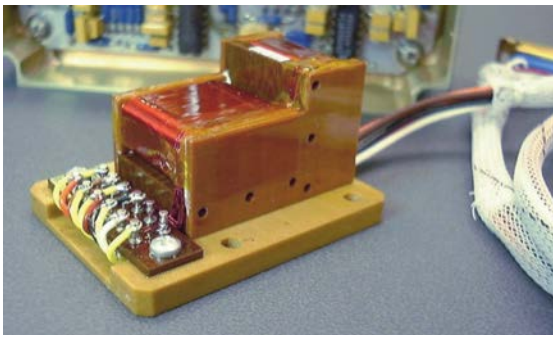


Figure 4.5.15: *Sensor head with two perpendicular "biaxial ring core fluxgate" elements. Each element allows for the measurement of two magnetic field vector components (Source: SAIC/Nanotesla).*

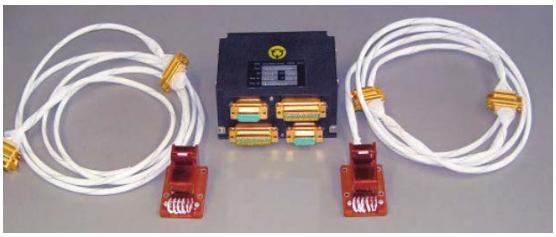


Figure 4.5.16: *Redundant magnetometer, consisting of an internally redundant electronic box and two sensor heads (Source: OHB-System).*

without specific postcalibration mechanisms. Special attention has to be paid to the fact that the measurement can be affected by offset components resulting from temperature impacts and the residual magnetic field of the satellite itself.

In general the sensing part of the magnetometer (Figure 4.5.16) is deployed as a **sensor head** on a boom away from the satellite, mainly for the following reasons:

- The measurement is less influenced by the residual static magnetic field of the satellite. This is even more valid for changing magnetic field components resulting from currents within the satellite or from payload activities.
- The electronic box of the magnetometer can be allocated to a free position in the core of the satellite in order to allow for only moderate temperature influences.
- The risk of unwanted magnetization of the sensor head during integration, test and operation can be minimized.

Table 4.5.6: *Typical values for fluxgate magnetometers.*

Aspect	Value	Comments
Measurement accuracy of the magnetic field	0.5°	With adequate post calibration: 0.05°
Measurement bandwidth	40 Hz	Cut-off frequency for 50% of the effective value (see Section 4.5.6.1)
Power consumption	0.3 W	–
Mass	0.5 kg	Sensor head + cabling + electronic box
Application area	LEO	LEO = Low Earth Orbit

Magnetic torquers, which are often used in combination with magnetometers, generate a strong magnetic field influencing the magnetic field measurements, even if the sensor head is deployed on a boom. For this situation a pulsed operation of the magnetic torquer can help (e.g., 600 ms active, 400 ms passive), with the measurements being taken at the end of the passive time slots. Attention has to be drawn to the fact that the measurement process has to be fast in order to exclude residuals from the previous active phase. Therefore, it is important that the magnetometer has a broad measurement bandwidth (> 20 Hz; Table 4.5.6).

4.5.6.6 Gyroscopes

Gyros measure **rotations in an inertial reference frame**. The big advantage with respect to other attitude sensors is the complete independence from external sources. The gyro signal is continuous and available to very good resolution. Gyros are therefore especially useful for stabilizing spacecraft rotation and for bridging time spans without direct attitude measurements.

The measurement value is provided either as the angular increment of the last measurement span or as the actual angular velocity. Gyro base attitude determination needs an initial attitude value. This value can then be propagated using the gyro measurements. The attitude can also be calculated as a mixture of direct attitude measurements and gyro values, the so-called "blended solution."

Gyro measurements are affected by a drift which can be described to a first approximation by the following error categories:

- **Bias:** The bias is the measured value for an inertial rotation rate of 0°/s. This error has a quasi-static component which changes over a range of days or between on–off–on switching cycles, as well as through contributions from external influences (e.g., magnetic fields or accelerations). Temperature has the main impact on the bias. Therefore, gyros have an internal temperature regulation. They should be mounted within the satellite at a position with a small local and temporal temperature gradient. The bias stability is a major quality criterion for a gyro.
- **Scale Factor Error:** The scale factor error describes (after subtraction of the bias component) the relative deviation of the actual rotation rate with respect to the provided rotation rate measurement. This error component is indicated in parts per million (PPM). The scale factor error becomes important if wide angular ranges are propagated without intermediate direct attitude measurements. As with the bias, this error component is also mainly influenced by the temperature.
- **Noise:** The random noise is in general very low for gyros. Nevertheless it plays an important role if various gyro increments are integrated for the attitude determination process. The sum of these errors forms the so-called "random walk," whose uncertainty increases in proportion to the square root of the number of measurements (or with time).

Constant or slowly varying error contributions are often determined by continuous reference measurements with drift-free attitude sensors on-board the satellite, and can therefore be mostly compensated during signal processing.

Gyro Designs

The following gyro designs are used in spacecraft:

- Mechanical gyros (very common, decreasing)
- Hemispheric resonator gyros (rare, increasing)
- Ring laser gyros (very common, increasing)
- Fiber optic gyros (rare, increasing).

Mechanical Gyros

The rate measurement of a mechanical gyro is based on the **gyroscopic effect**: an axisymmetric rigid body is put into rotation and therefore has an angular momentum. The angular momentum is an inertial value, only influenced by torques:

$$\boldsymbol{M} = \dot{\boldsymbol{H}}_{\mathrm{g}} \tag{4.5.30}$$

where:
$\boldsymbol{M}$ = torque acting on the rotating mass,
$\dot{\boldsymbol{H}}_{\mathrm{g}}$ = angular momentum of the rotating mass.

As long as there are no **torques** acting, the angular momentum is constant with respect to size and amount. The angular momentum vector forms an inertial reference. With a corresponding torque-free gimbal mounting of the rotating mass, the attitude angles with respect to the angular momentum axis can be directly measured.

From a technical perspective it is less complex to align the angular momentum vector with the rotation and to measure the resulting torque.

For single axis measurement, the gyro consists of a rotor, elastically suspended in a frame. The frame can be tilted (see Figure 4.5.17). If the gyro is rotated around its measurement axis, **precession torque** is generated, which results in a declination of the angular momentum axis. This declination angle is proportional to the angular velocity.

The measurement accuracy can be increased by keeping this declination (deviation) continuously at

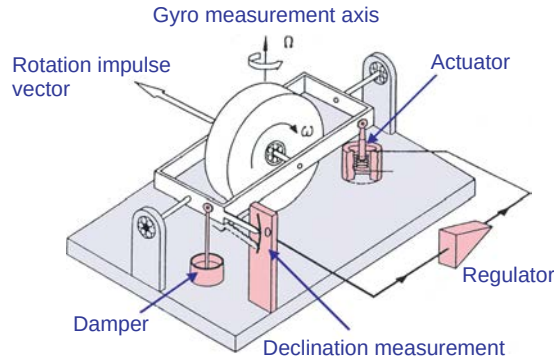


Figure 4.5.17: *Principle of rotation rate measurement for a mechanical gyro (Source: Lawrence).*

zero by using a controller. The generated compensation torque is proportional to the rotation rate to be measured.

Mechanical gyros (Figure 4.5.18) can achieve high accuracies (e.g., 0.001°/h, 3σ).

Oscillating rotation rate sensors are based on the measurement of variations of the oscillation plane of, for example, a ceramic bar which is made to oscillate by a piezoelectric actuator. They are almost never used in space applications because of their low accuracies.

A special position is held by the **hemispheric resonator** (HRG). The HRG uses (instead of a bar or fork) an open shell, comparable to a wine glass, which is made to oscillate. The generated oscillation form behaves inert, related to rotation. Resulting form shifts can be precisely detected and build a measure for the rotation of the sensor. HRGs are almost not subject to deterioration and aging processes. High accuracies, comparable to mechanical or laser gyros, can be achieved.



Figure 4.5.18: *Technical design of a mechanical gyro (Source: Kearfott).*

Fiber Optic Gyro (FOG)

An FOG is a **passive laser gyro** consisting of a coil made from monomode optical glass fibers, into which the laser light is coupled from the outside (Figure 4.5.19).

The laser light is divided into two beams by a semipermeable mirror and guided into different ends of the same glass fiber. The beams flow into the fiber in opposite directions. With a rotation of the gyro slightly different times are needed for the two beams. Both beams are overlaid again by the semipermeable mirror. The phase difference resulting from the angular velocity causes an amplification or attenuation of the overlaid beam light.

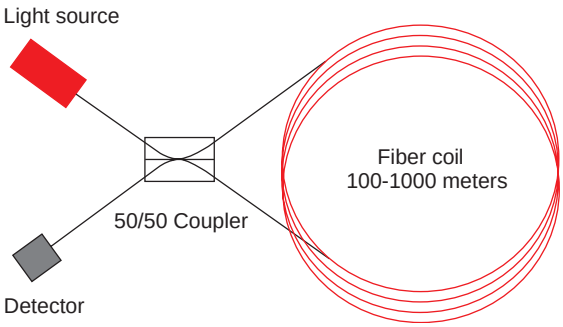


Figure 4.5.19: *Principle of the fiber optic gyro (Source: Teldix).*

FOGs are lightweight, have low power consumption and very high accuracy. However, in the past the degradation of the gyro's optical elements has limited their use on satellites.

Ring Laser Gyro (RLG)

A closed optical path is formed by (e.g., three) mirrors, in which a laser active gas is contained (Figures 4.5.20 and 4.5.21). After activation, two opposite running

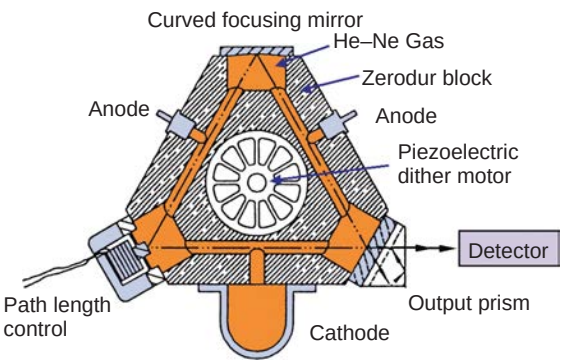


Figure 4.5.20: *Function scheme of a ring laser gyro (Source: Lawrence).*

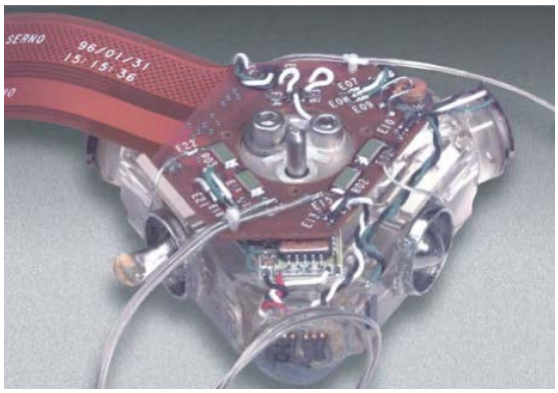


Figure 4.5.21: *Ring laser gyro glass block (approx. 5 cm) with electronic cabling (Source: Honeywell).*

beams are generated, which are overlaid to a standing wave. This wave forms an inertial reference. During rotation – from the perspective of the gyro – the beam running in the antirotational direction seems to have a slightly reduced wavelength, whereas the beam running in the rotational direction seems to have a slightly higher wavelength. By partial decoupling and superposition of both light beams a beat can be measured at the detector, whose frequency is proportional to the rotation rate.

Problems

Qualitatively, high-class gyros require a very complex manufacturing and calibration process; they are therefore very expensive and sensitive measurement devices.

For long-term missions, gyros have often been the reason for anomalies and failures. Therefore, ways to reduce their use on spacecraft or even avoid them (gyro-less spacecraft) are under consideration. In practice, solely with star sensors (and specific attitude filters) various attitude processes can be carried out, but there is almost no alternative to gyros in case of demanding requirements for availability, stability and agility (Table 4.5.7).

4.5.6.7 Global Navigation Satellite Systems for Attitude Determination

Although GNSSs like GPS, Galileo or GLONASS are primarily designed for position determination, in principle it is possible to also determine the attitude with a GNSS receiver. This is shown in the following, using GPS as an example.

Attitude determination with GPS is based on an evaluation of GPS signals received simultaneously at different positions. The distance between two receiving antennas is called the baseline. The carrier phase of the GPS signal is received at the antennas with slightly different reception times. The resulting phase difference can be measured; it is the cosine of the angle between the antenna baseline and the direction of the GPS satellite transmitting the signal (Figure 4.5.22).

By combining phase difference measurements with different GPS satellites, the direction of the baseline in space can be completely determined – and thus, using two linear independent baselines, the three-axis attitude of the satellite. For this process at least three antennas are needed, which are positioned as shown in Figure 4.5.23.

Comparison of the carrier phase differences assumes a simultaneously available and authentic signal at all receiving GPS antennas of the satellite. For the satellite design it is important that these antennas have a similar field of view and that the signal cannot be disturbed by

Table 4.5.7: *Typical values for gyros.*

Aspect	Mechanical gyro	HRG	RLG	FOG	Comments
Bias	0.01°/h	0.01°/h	0.01°/h	1°/h	1σ, after compensation
Scale factor error	50 PPM	30 PPM	100 PPM	1000 PPM	1σ, after compensation
Random walk	$0.005°/h^{1/2}$	$0.0005°/h^{1/2}$	$0.01°/h^{1/2}$	$0.3°/h^{1/2}$	1σ
Degradation	Deterioration	No	Aging	Yes	–
Moving parts	Yes	No (oscillating)	No (dither)	No	–

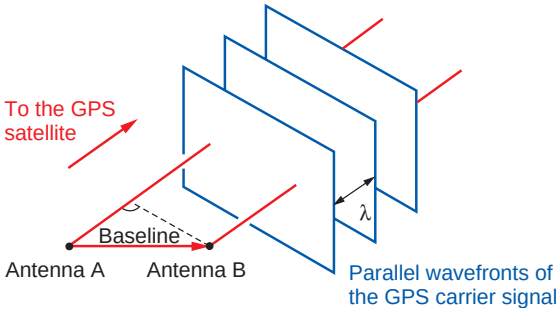


Figure 4.5.22: *Principle of GPS-based attitude determination (Source: R. Bock).*

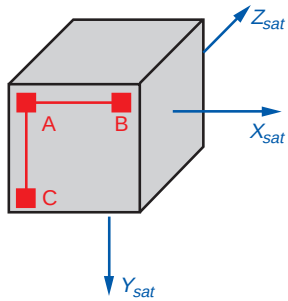


Figure 4.5.23: *Positioning of three GPS antennas A, B, C on a satellite (Source: R. Bock).*

shadowing effects (satellite structure) and reflections (**multipath effects**).

The achievable accuracies of GPS-based attitude determination depend fundamentally on the baseline length (the longer, the better). The actual attitude accuracy varies with the geometry of the visible GNSS constellation.

The attitude process is started after resolving ambiguities which are inherent for GNSS-based attitude determination. These ambiguities are resolved in the processor of the GPS receiver by using information on the actual satellite position and the positions of the GPS satellites.

The use of GNSS signals for attitude determination is a quite new discipline and can be implemented easily from a technical perspective. There is high potential with respect to achievable performance values and possible application areas (Table 4.5.8). This is also valid for the Galileo system and corresponding receivers.

Table 4.5.8: *Typical values for GNSS-based attitude determination.*

Aspect	Value	Comments
Attitude accuracy	0.1°	1.5 m baseline, stationary measurement setup
Power consumption	7 W	Inclusive position determination
Mass	3 kg	Inclusive position determination Single receiver, four antennas with LNA and harness
Application area	LEO	Studies: use in GEO is possible

4.5.7 Actuators for Attitude Control

4.5.7.1 General Aspects for Actuators

Spacecraft attitude can only be influenced by torques.

A torque does not adjust the attitude directly, but causes a change in the angular momentum, which is proportional to the angular velocity of the spacecraft. The following formula applies to the simplified one-axis case:

$$\ddot{\varphi} = \frac{T}{I} \tag{4.5.31}$$

where:
T = torque,
I = moment of inertia (constant for a rigid body).

The attitude, represented by the angle φ, results after two integrations. Thus a device that generates a torque in a defined way can be used as an actuator for attitude control.

External and Internal Torques

External torques result in interaction with the spacecraft environment. The usable torque has an effect on the satellite attitude; its corresponding reaction torque remains outside the spacecraft system and needs no further consideration. External torques can be generated, for example, by thrusters or magnet torquers.

An internal torque affects the satellite attitude in the same way, but its corresponding reaction torque remains within the spacecraft system and has to be

considered further. For example, a **reaction wheel** accelerates its flywheel and thus generates a torque in the spacecraft. The angular momentum of the flywheel rotation speed change is taken over conversely by the rotation of the spacecraft structure; the overall angular momentum of the spacecraft system remains constant. Thus internal torques can only internally distribute parts of the spacecraft's angular momentum, within the capacities of its actuator elements.

Since external disturbance torques can be compensated by internal torques only for a limited duration, "momentum management" with external torques has to be regularly undertaken. In LEO external disturbance torques always exist for satellites with constant attitude. Continuous attitude control without involving any external actuator torque is therefore not possible.

4

Required Properties of Actuators

For a three-axis attitude control process the accuracy requirements for actuators are not as stringent as for attitude sensors. Small-scale errors result in slightly changed loop characteristics and can be generally afforded. Small errors in alignment or biases of actuators are compensated inherently by the three-axis control process, but have to be considered specifically for control about only one rotation axis.

More important are **nonlinear properties**, such as quantization, jumps or switching. As noncompensatable disturbances they affect and limit the achievable stability of control (Table 4.5.9).

Table 4.5.9: *Characteristics of actuators for attitude control.*

Actuator	Torques	Remarks
Thrusters	Small to very high torques, quantized	External torque, needs fuel
Magnet torquer	Several mNm, only in two axes	External torque, cost effective
Reaction wheels	≈ 200 mNm, variable	Internal torque
Momentum wheels	≈ 50 mNm, variable, momentum bias	Internal torque, momentum stabilization
Control momentum gyros	Temporarily very high torques	Internal torque, expensive

Furthermore, possible side effects of actuators on the satellite system, such as magnetic fields or vibration, have to be considered.

Further Actuator Elements

The following elements can be regarded in the broader sense as attitude actuators. They either provide control torques passively, or under specific conditions transform disturbance torques into control torques.

Nutation Damper

Nutation dampers are used as passive actuators for spin-stabilized satellites. Such a damper can be designed simply, for example as a movable spherical mass in a closed tube filled with a fluid. Nutation causes movement of the mass. The damping fluid takes up a part of the surplus rotation energy and therefore reduces nutation.

Gravity Gradient Boom

A boom separates a tip mass from the satellite's center of gravity. A body with such a long-shaped boom is subject to the (Earth's) gravity gradient and creates a torque which causes the spacecraft's longitudinal axis to align radially to the Earth's center.

Gravity gradient stabilization is a simple method to achieve coarse nadir pointing for small satellites in LEO. Nevertheless, this stabilization should be supported by controlled damping with magnet torquers.

Mechanisms, Sun Sailing

Sunlight causes a slight pressure on the irradiated surfaces of a spacecraft (e.g., the solar panels). The resulting forces and corresponding torques can be varied with a slightly altered steering of the solar panels and can therefore be regarded as small control torques. This kind of Sun sailing is used by geostationary satellites for momentum management.

4.5.7.2 Reaction Wheels, Momentum Wheels, Control Momentum Gyros

Reaction wheels provide a reaction torque which results from rotational acceleration (or deceleration) of a flywheel. The flywheel rotation is driven by an electric motor, which allows a variable adjustment of torque.

The torque practically usable for attitude control is opposite to the flywheel's angular momentum change and follows the equation

$$\boldsymbol{M} = -\dot{\boldsymbol{H}}_R = -I_R \cdot \dot{\boldsymbol{\omega}}_R \qquad (4.5.32)$$

where:
$\boldsymbol{H}_R$ = angular momentum vector of the flywheel,
I_R = moment of inertia (about the flywheel rotation axis),
$\dot{\boldsymbol{\omega}}_R$ = angular velocity vector of the flywheel.

Three cases can be distinguished:

1. The rotation speed of the flywheel is low. Acceleration or deceleration causes a reaction torque aligned with the flywheel axis. Wheels designed for this operational mode are called **reaction wheels**.
2. The rotation speed of the flywheel is high, and so is, accordingly, its angular momentum. A torque transverse to the rotation axis causes a slow precession (movement) of the wheel, which compensates the transverse torque. This gyroscopic effect achieves stabilization for the transverse axes of the wheel. Wheels which are designed for this purpose are called **momentum wheels**.
3. A momentum wheel can be suspended in a gimbal that can force tilting of the spin axis. The tilting causes (conversely to gyroscopic stabilization) very high reaction torques. A gimbaled wheel built for this purpose is called a **control momentum gyro** (CMG).

Disturbances

In reaction and momentum wheels the flywheel mass is commonly supported by robust ball bearings which can sustain the launch loads but have the following side effects as well:

1. Microvibration, caused by the rolling of the balls and by the movement of the bearing cage.
2. Vibrations caused by static and dynamic imbalances of the flywheel (typically $<1\,\mathrm{g\,cm}$, or $<10\,\mathrm{g\,cm^2}$). Amplitude and frequency of these vibrations increase with rotation speed.
3. Reversal of friction torque at wheel speed zero crossing (typically $\pm 10\,\mathrm{mNm}$) and variation of friction (e.g. 1 mNm) at steady-state speed.

Reaction Wheels and Momentum Wheels

The concepts of reaction and momentum wheels are in principle the same; the differences are in the dimensioning of the motor and flywheel. Reaction wheel motors are designed for high torque that is provided at low rotation speeds. The flywheel, running at different speed levels, is commonly manufactured as an integral part with low variation of imbalances over its full speed range. Momentum wheels, running constantly at high speed, are equipped with a spoke wheel that is more mass effective and provides better structural damping.

The housing (of both types) is generally evacuated to a low pressure level which avoids on the one hand the vaporizing of the lubricant in the bearings, and on the other hand limits the atmospheric friction.

The electronics needed to drive the motor are usually placed inside the housing, but may also be separated for better heat dissipation (Figure 4.5.24).

Off-the-shelf reaction and momentum wheels are available for angular momentum capacities from 0.1 to 250 N ms.

Control Momentum Gyro

A control momentum gyro (CMG) is a gimbal-mounted momentum wheel (Figure 4.5.25). A distinction is made between **single-gimbal CMGs**, which can be rotated around a single transverse axis, and **dual-gimbal CMGs**, which can be turned around two transverse axes like a Cardan joint.

Turning transverse to the angular momentum results in torques which may be 100 times higher than the usual reaction torques of a reaction wheel. This allows substantial improvement in satellite pointing agility.

The generated torques change their direction with the orientation of the flywheel. That is why the high torque is only available within certain boundaries and why singularities occur even if CMGs are working together in a cluster arrangement.

For complete three-axis attitude control at least three single-gimbal GMGs are needed.

Compared to reaction wheels, CMG technology requires much more volume and mass. Therefore, CMGs are preferably used in large satellites.

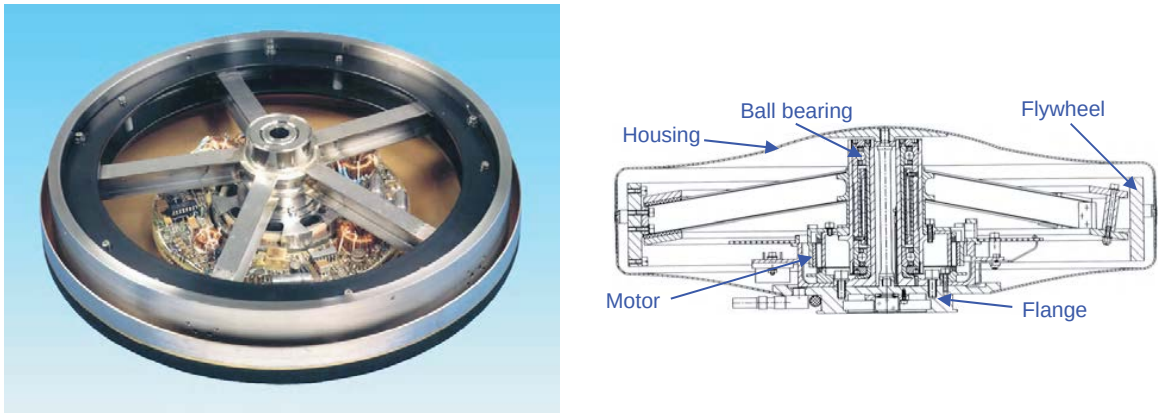


Figure 4.5.24: *View from above and schematic cross-section of a momentum wheel (Source: Teldix, Rockwell Collins).*



Figure 4.5.25: *Control momentum gyro (Source: Honeywell).*

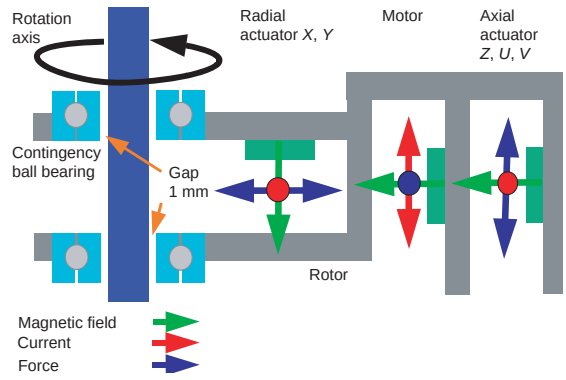


Figure 4.5.26: *Schematic view of the magnetic forces of a five-axis controlled magnetic bearing wheel (Source: Teldix, Rockwell Collins).*

Magnetic Bearing Wheel

The magnetic bearing wheel uses only magnetic forces to suspend the flywheel in and around all its axes (Figure 4.5.26). The flywheel's position and angular definition are actively controlled by a processor which allows soft and adjustable transfer of forces and torques, providing the following advantages:

1. Reduced microvibrations (by a factor of 10) compared to ball bearing reaction wheels.
2. No bearing friction and therefore no variation of bearing friction over the full speed range. This allows more precise control of satellite slew maneuvers.
3. Magnetic bearings provide the possibility to generate (over the short term) strong tilting torques.

For present concepts the overall mass and required electrical power of a magnetic bearing wheel are very high compared to ball bearing reaction wheels (Table 4.5.10). Thus magnetic bearing wheels are mainly considered for missions which require extremely low vibration disturbance levels.

Table 4.5.10: *Characteristic values.*

Actuator	Reaction wheel	Momentum wheel	Control momentum gyro	Remarks
Momentum capacity	15 Nm s*	50 Nm s	20 Nm s	At nominal speed; *at full speed
Reaction torque	200 mN m	50 mN m	45 000 mN m	Maximum
Mass	7 kg	8 kg	15 kg	––
Power	10 W	15 W	25 W	At steady-state speed. The temporary power consumption at full reaction torque is much higher

4.5.7.3 Thrusters for Attitude Control

Thrusters are commonly the only possibility to generate sufficient external control torque at high altitudes or in outer space. They are applied also for large torques, as needed for attitude control during orbit control maneuvers.

The usable torque results from a combination of thrust force and lever arm. Torque levels from 0.1 to 10 Nm or more can be easily achieved. However, the following aspects have to be regarded.

Fuel

Thrusters need fuel, which may be supplied from an existing on-board orbit control propulsion system. Since attitude control thrusters are small, their temporary use for momentum management or wheel unloading does not consume a significant amount of fuel. On the other hand, continuous and precise attitude control over years, using thrusters alone, would not be reasonable.

Cold Gas as Fuel

For scientific missions, users frequently require the a priori exclusion of contamination of instruments or disturbance of the magnetic field. These are usually short- or mid-term missions for which *post facto* attitude information is relevant. For such cases, attitude regulation using cold gas is an option (Figure 4.5.27).

Impulse Bit

Switching thrusters on or off requires a certain duration that results in a minimum thrust effect, the so-called impulse bit. The size of this impulse bit may limit the stability of attitude control achievable with thrusters.

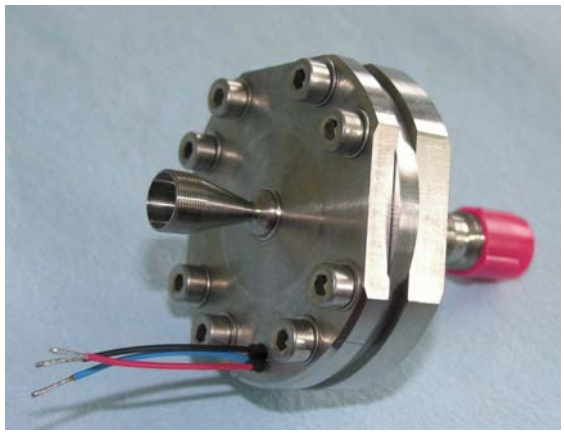


Figure 4.5.27: *Cold gas thruster (with piezovalve) of 40 mm diameter (Source: RTG Aero-Hydraulic Inc.).*

Accuracy of Attitude Control

Continuous operation at high control accuracy (without leaving a "dead band" around the target) would require permanent alternation between "negative" and "positive" thruster activation. Besides incessant fuel consumption, the high "pulse rate" could excite oscillations in large satellite structures, such as solar panels or antennas.

Number of Thrusters

Thrusters cause primary forces. In order to avoid any side effects from attitude control in the spacecraft's orbit, attitude control thrusters should be operated in pairs, separated by a lever arm and firing in opposite directions.

Table 4.5.11: *Typical characteristics of thrusters for attitude control.*

Thruster	Hydrazine monopropellant	Cold gas	Remark
Thrust	1 N	0.1 N	–
Flow of mass	450 mg/s	130 mg/s	At nominal thrust
"Impulse bit"	0.3 N m s	0.001 N m s	For 1 m lever arm
Power consumption	20 W	10 W	Hydrazine monopropellant: heating up required
Mass	0.4 kg	0.1 kg	Without piping and valves

4

Handling and Operation

For attitude control thrusters, mainly hydrazine (with or without an oxidizer) is used (see Table 4.5.11). This fuel, generally used for spacecraft orbital control, is toxic. Thus, testing attitude thrusters under laboratory conditions is not practical.

4.5.7.4 Magnetic Torquer

A magnetic torquer (also referred to as a "torque rod") generates a strong **magnetic dipole** that interacts with the Earth's magnetic field. Thus the torquer is subject to an external torque, like a compass needle.

The dipole can be generated by a simple coil (of copper or aluminum); its strength is proportional to the area which is enclosed by the electric current and to the current itself. Mass-effective designs can be made in this simple way up to dipoles of 10 Am^2. To generate stronger dipoles, the coil is wrapped around a core material, typically a rod of soft iron that amplifies the magnetic flux inside the coil. These "torque rods" (Figure 4.5.28) are designed for typically 50 to 250 Am^2 dipole moments.

Usable Control Torque

The torque which can be used for the attitude control process is proportional to the dipole moment of the torquer and the strength of the Earth's magnetic field and depends on the angle between these two.

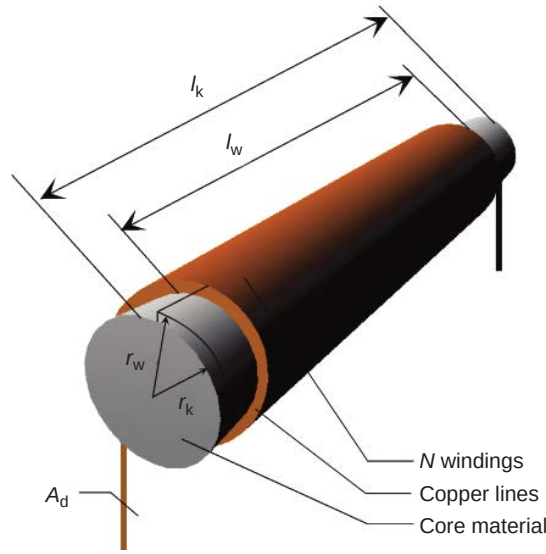


Figure 4.5.28: *Design of a torque rod (Source: ZARM, Bremen).*

It can be formulated in a straightforward way as a cross-product:

$$\boldsymbol{T}_q = \boldsymbol{D} \times \boldsymbol{B} \tag{4.5.33}$$

where:
$\boldsymbol{T}_q$ = established torque vector [N m],
$\boldsymbol{D}$ = dipole moment vector [A m^2],
$\boldsymbol{B}$ = Earth's magnetic field strength vector [T].

Example: A torque rod with a dipole of 50 A m^2 at the equator (B = 20 μT to the north) that points east generates a torque of 1000 μN m, pointing to the zenith.

Applications

Magnetic torquers generate quite weak external torques only in a plane perpendicular to the Earth's magnetic field vector. In polar orbits, where the Earth's magnetic field changes its orientation continually, all torque directions can be generated within a reasonable period of time with a set of magnetic torquers pointing in different directions. Thus torquers are suitable for continuous angular **momentum management** or for **wheel unloading** in highly inclined LEOs.

Because of their simplicity, magnetic torquers (Figure 4.5.29) are cost efficient and reliable. In contrast to thrusters, only electric current is needed for operation (Table 4.5.12).



Figure 4.5.29: *Magnetic torquer (250 Am^2) with soft-iron core (Source: ZARM, Bremen).*

Table 4.5.12: *Typical values for magnetic torquer.*

Design	Coil (e.g., 0.3 × 0.3 m^2)	Torque rod (0.85 m)	Remarks
Typical dipole moment	5 Am^2	200/220 Am^2	Coil with iron core: linear/maximal
Usable torque	75–250 µNm	3000–10 000 µNm	Nominal, in LEO
Remanence	–	<0.5%	Remaining dipole, after switch-off
Mass	0.8 kg	5 kg	–
Power consumption	1.5 W	3 W	At 28 V, 100% duty cycle
Typical application	Satellites < 50 kg	Satellites > 100 kg	In polar and LEOs

4.5.8 Verification of the Attitude Control System

The ACS is one of the most complex systems on-board a satellite. It consists of various components, such as several sensors and actuators, their corresponding mechanical and thermal interfaces to the satellite bus, electrical and data interfaces with the board computer, the software for the attitude control process, and operational interfaces to the ground segment. Moreover, the ACS is constructed with backups to minimize the probability of single point failures.

All these elements, functions and the system itself have to be verified within the qualification process (which verifies the suitability of a new design) or within the acceptance process, (which demonstrates that the implementation of the design and the manufacture have been carried out properly).

In Europe the formal verification steps themselves and the standards to carry out each of these steps have been defined by the European Cooperation on Space Standardization (ECSS), which was established by ESA.

It is common for the main functions of the satellite, particularly the processes of the ACS, to be regularly checked by a **software module** which detects critical states and automatically switches over to alternative modes or to redundant hardware. The module goes by the name of **fault detection, isolation and recovery** (FDIR). Its autonomous functions may essentially affect the attitude control and have to be verified carefully at the system level as well.

Furthermore, the ACS elements may interfere magnetically or electrically with other satellite subsystems or even with each other (e.g., torque rod activation may disturb magnetometer measurements). These aspects are summarized by the terms **EMC (Electromagnetic Compatibility)** or **EMI (Electromagnetic Interference)**.

Methods for testing and verification can also be seen as instruments for further development; they should therefore begin as early as possible. Table 4.5.13 summarizes the most important verification steps and their main aspects.

Since the conditions of space are not present on Earth, tests become difficult as soon as physical

Table 4.5.13: *Aspects of verification of an attitude control system.*

Verification	Aspects	Methods
ACS elements hardware (e.g., sensors)	Specified performances	Tests, carried out by the element supplier
	Functions on the satellite	Test after satellite integration, qualitatively
Hardware interfaces	Axis assignment, alignment	Measurement on the spacecraft
	Axis polarity	Tests on the spacecraft
	Thermal	Analyses and thermal vacuum test
	Electrical	Tests and measurement on the spacecraft
EMC, EMI	Between ACS elements	Tests and measurement on the spacecraft
	Between ACS and satellite (emission, susceptibility)	Tests and measurement on the spacecraft
	Possible hardware configurations (redundancy)	Tests and measurement on the spacecraft
Software interfaces	Formats, timing, conversions	Software simulator, tests on the spacecraft
	ACS telemetry and telecommand	Software simulator
Attitude control processes		
Operations	Operational sequences	Software simulator, HILT
FDIR	Fault detection, Corrective measures	Software simulator, HILT
Attitude control	Stability of processes	Analyses, software simulator, HILT, test bed
	Control deviations	Analyses, software simulator, HILT, test bed
	Course and duration of processes	(Analyses), software simulator, HILT, test bed

properties and their connected control processes are involved. On the other hand, exactly these processes and their characteristic values, such as natural frequencies, stability, control deviation, etc., are of particular interest and have to be specifically determined.

These properties can be verified using a graduated approach:

- Analytically
- By software simulation
- By HILT (Hardware In the Loop Testing)
- In specific test beds, for example on an air bearing.

4.5.8.1 Analytical Verification

Each element of the attitude control loop (sensors, controller, actuators and plant) can be represented by a **physical model** with a mathematical equation as transfer function. The equations should be linear and kept as simple as possible. The closed chain of these transfer functions represents the attitude control loop, commonly as a linear differential equation. For this type of equation a solution always exists and can be found elegantly by applying Laplace transforms.

The advantage of such an analytical solution is its closed form, which allows immediate assessment of the stability of the control loop and easy optimization of the controller parameter.

On the other hand, state-of-the-art controllers are based on computer programs working with complex and nonlinear processes. Simplification and linearization of these processes are in most cases not possible without inadmissible distortion of the mathematical models.

4.5.8.2 Software Simulation

Verification of the attitude control processes can also be performed purely by software simulation. In this case the different elements of the attitude control loop (sensors, actuators, environment, etc.) are represented by software modules. These modules copy (as authentically as possible) the element characteristics and can even reflect nonlinear and complex processes such as conditioned mode transitions, switching and

algorithms. In particular the software for attitude determination and attitude control can be embedded as identical modules.

The complete attitude control loop is represented by linking its several element modules to form a closed chain.

The simulation itself is done numerically and provides a particular solution for a specific case. The general behavior of the attitude control loop can be established by running a series of simulations with different settings for module parameters and initial conditions. Parameter variation can be done systematically or randomly by **Monte Carlo simulations**.

Simulation is a commonly used method to verify existing concepts or developments (Figure 4.5.30). Simulation tools such as MATLAB/Simulink allow in turn the design of control loops and automatically generate the executable software code from symbolic modules which can be directly embedded in the on-board computer software. This method reduces both development costs and the probability of implementation errors, particularly for large and complex software.

4.5.8.3 Hardware in the Loop Test

For HILT (see Figure 4.5.31), as many elements as possible of real attitude hardware are used in the attitude control loop. This is particularly true for the on-board computer, using its original interface cards

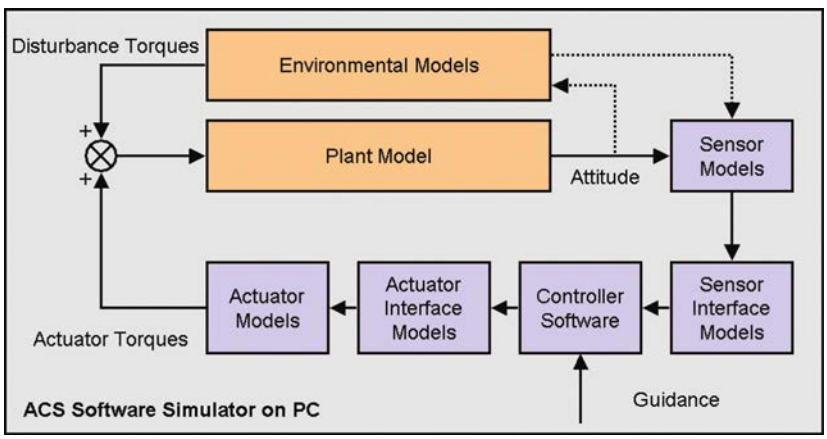


Figure 4.5.30: *Software simulation for the attitude control loop.*

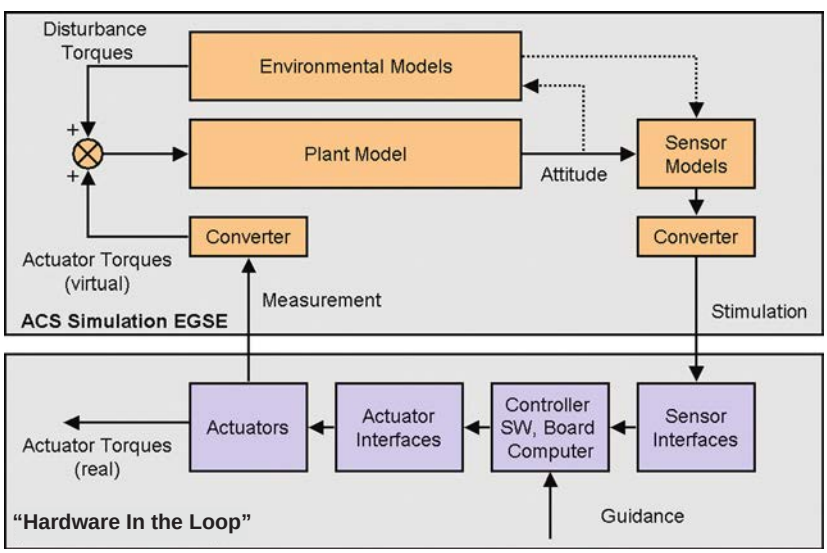


Figure 4.5.31: *"Hardware in the loop" testing.*

and control software. HILT runs in real time and is therefore practical for verification of the implemented attitude control subsystem.

An important aspect for verification is the separate development of on-board data handling (OBDH) control software and the "verification simulator". Such a verification simulator can be suitably developed by another company or a separated in-house department. Strict separation of both developments reduces the risk of double errors (e.g., in axis polarity, scaling, etc.) which are critical but compensate each other in the test setup and thus would not become visible during verification.

4.5.8.4 Air Bearing Test Bed

Even more realistic than HILT is end-to-end verification of the completely integrated ACS on an air bearing test bed. But this is subject to restrictions:

1. For sensors such as star trackers or Earth sensors, the environmental conditions of space and their dependence on the current attitude cannot be duplicated; for magnetometer or Sun sensors they can only be approached. Attitude propagation using gyros is possible without restriction. Thus the closed attitude control loop can only partially be tested with original sensors.
2. Disturbance torques, present in space, are in the range of a few millinewton meters (for small satellites, 100 µNm). Test beds are subject to higher disturbances, even if they are ideally suspended. Hence an attitude control using weak actuators can be verified only qualitatively. Furthermore, the center of gravity of the elastic test bed table moves slightly under the influence of gravity, which limits the angular corridor for authentic testing. Geometry also restricts it.
3. Spacecraft rotation axes go through the center of gravity, which is normally located inside the original satellite shape. Air bearing tests are therefore carried out with a special satellite engineering model containing a representative arrangement of attitude control elements. Air bearing verification with the completed original satellite flight model is usually not possible.

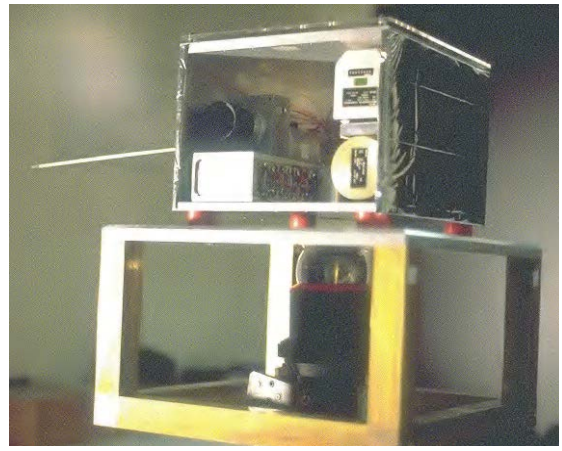


Figure 4.5.32: *Air bearing test bed at TU Berlin, 1994 (Source: TU Berlin).*

In certain cases, such as for the qualification of new attitude control methods or HILT of control momentum gyros, the above-mentioned restrictions play a minor role.

Example 1
The test setup is put on a sphere (in the center of Figure 4.5.32), which is supported in a shell by a continuous flow of pressurized air. Thus, the whole table can rotate freely. The satellite engineering model (on the table) works independently; it is equipped with its ACS, batteries and a telemetry and telecommand unit. The total rotating mass is about 20 kg.

The massive frame in the lower part defines the center of gravity in the center of the sphere. The rigidity of this setup and its geometry allow limited slewing around the horizontal axes and unlimited rotation around the vertical axis.

Example 2
In this test setup the model of a complete (micro)satellite is integrated into a spherical shell of about 30 cm diameter. This small satellite model contains all the subsystems needed for independent operation and for demonstrating the attitude control processes (i.e., attitude control elements including small cold gas thrusters, the electrical power subsystem, telemetry and telecommand, on-board computer).

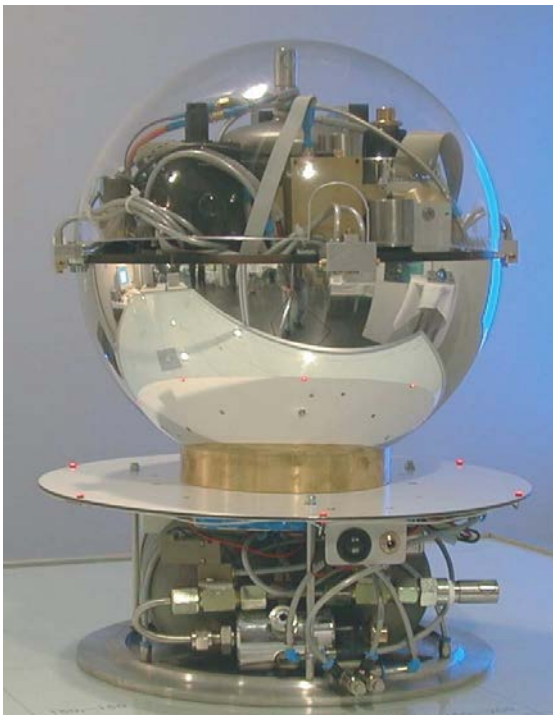


Figure 4.5.33: *Air bearing test bed at IAT Bremen, 2005 (Source: IAT Bremen).*

Furthermore, there is an additional air suspension on the flat table, which allows lateral gliding. Hence the satellite model can rotate freely around its three axes and be moved by small, cold gas thrusters along its horizontal axes. This small air bearing test bed shown in Figure 4.5.33 is a development of Institute for Aerospace Technology (IAT), Bremen in cooperation with EADS.

Bibliography

[4.5.1] Wertz, J.R. *Spacecraft Attitude Determination and Control.* Dordrecht: Kluwer Academic, 1997.

[4.5.2] Wertz, J.R. *Mission Geometry: Orbit and Constellation Design and Management.* Dordrecht Kluwer Academic, 2001.

[4.5.3] Thomson, W.T. *Introduction to Space Dynamics.* New York: Dover, 1986.

[4.5.4] Sidi, M.J. *Spacecraft Dynamics & Control.* Cambridge: Cambridge &University Press, 1997.

4.6 Data Management

Sergio Montenegro

4.6.1 Data and Information Management On-board

Basically (almost) every spacecraft serves "only" as a data capture and information collection center for the world. In this sense a spacecraft is nothing more than a data processing machine. Data is registered by the sensors, processed by the on-board computers, stored, and then transmitted by telemetry to the ground. Hence, data management is the central element of a spacecraft. The data management segment controls all other devices and collects the provided data. Information is extracted from this data and then transmitted to the ground.

In principle there are two different **data management segments** on-board with two different functions (see Figure 4.6.1). The first segment's task is **control and surveillance**, traditionally executed by the **on-board computer**, which controls the spacecraft so that it survives and fulfills its mission. The second segment's task is **payload monitoring and control**. This function is traditionally executed by the **payload computer**, which collects the payload data and information; that is, it supports the scientific mission of the spacecraft. We say "traditionally" because at the moment

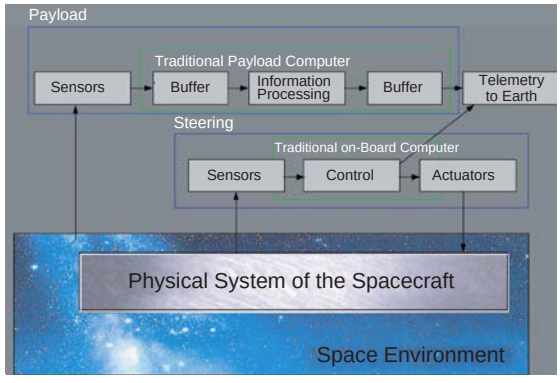


Figure 4.6.1: *Traditional approach to the structuring of data management in the space segment.*

there are two trends: an attempt to execute everything on a powerful, dependable central computer, versus an attempt to distribute all functions over many smaller computers.

4.6.1.1 Differences to "Normal" Earth (Ground) Devices

The data management system on-board is similar to many other embedded systems on the ground; but in space we have very strict constraints and difficulties, as follows

Special Requirements for Space Applications

- **Reliability:** This property is extremely important for the data management system and is the main cause for its high expense in comparison to terrestrially embedded systems (see Section 4.6.4.). Three factors determine reliability: robustness, self-healing (self-regeneration) and fault tolerance.
- **Self-healing (self-regeneration):** The spacecraft must be able to handle and treat failures and anomalies by reconfiguration using redundant resources (reserves and spares). This is especially important for missions with long lifetimes, for example 15 years. In these cases, simple self-healing using cheap components is not enough. Here we rather need very robust components. This originates a trade-off: how much to invest in **robustness** to avoid failures, and how much to invest in **redundancy** to compensate for failures?
- **Fault Tolerance:** Even if components are not permanently damaged, malfunctions are to be expected at any time. The system must be able to recognize such anomalies, to compensate and to correct them, favorably, before they have unexpected consequences. On Earth we find similar requirements (and even stricter ones) in the case of **safety-critical systems**, such as for railway and aircraft transportation systems or for nuclear reactor control. Nevertheless, on the ground we are well protected by an approximately 50 km thick atmospheric layer against cosmic radiation; a spacecraft is, however, not protected. This radiation causes a huge amount of **data corruption** (**bit flips**) and leads to much faster **aging** of the electronic components. Exactly as with terrestrial safety-critical systems, data corruption must be treated. However, its frequency is higher in space by approximately a factor of 10 up to 1000; 1 to 10 **bit flips** per day can be expected to occur in the on-board computer memory (e.g. 16 Mbytes).
- **Very Limited Space, Energy (Power) and Dimension Resources:** As with hand-held devices (cellular phones, portable navigation systems, MP3 players, organizers, cameras) the system has to use those resources very economically. Solutions for such Earth products can be of interest also for space missions if they are able to survive space conditions.
- **Cold/Heat:** Although the spacecraft is exposed to extreme temperatures, from −170 to +120 °C, the temperature range of the data management segment is relatively harmless. The data management segment is rather well protected in the middle of the spacecraft. The temperature mostly stays between +10 and +40 °C, which produces no difficulties for the electronics. With some terrestrial applications the requirements are more extreme: for instance, automotive electronics must be able to work from −40 (winter nights) to +80 ° C (direct Sun in summer).
- **Vacuum:** For electronics on the ground, we can use air (even ventilators) to conduct heat away from the heat source. In space, in vacuum, it is not so simple to conduct heat away. This can lead to a heat concentration and a very irregular temperature distribution in the electronic components. Metal heat conductors must be used to transport heat from the electronic boards to the main spacecraft structure.
- **No Gravity:** This should be no problem for the electronics and the software, but a loose, small metal part (e.g., a cut-off piece of wire) could freely move around and cause short circuits at different places. We have similar problems in terrestrial applications experiencing strong vibrations, for example in moving vehicles.
- **Vibrations:** In free flight in orbit vibrations are hardly to be expected, but during the launch, for about 10 to 15 minutes there are extreme vibrations on the rocket (launcher) up to seven times the

Earth's gravitational acceleration (g). At separation there is a shockwave of a few microseconds at up to 1000g. This requires especially robust hardware.

- **Software Complexity:** As in all other information technology areas, software complexity increases rapidly. It has now reached dimensions which are barely controllable. With today's complexity one must use proven software engineering methodology and a strict quality assurance program. The expenses for software quality assurance are often underestimated; they amount to between 50 and 80% of the complete software development cost.
- **Software Uploads:** Software efforts are almost always underestimated. Hence, some missions start before the software is ready. This is not a good concept and carries many dangers in itself. Nevertheless, it is important to create the capability of reloading new software or software updates. This functionality is also usually provided for mobile consumer electronics and navigation systems. Software uploads are necessary because it is not possible to foresee all possible situations on-board the spacecraft. During the mission improvements are necessary or errors might be recognized and new software must be installed on-board the spacecraft. Another reason is that in the course of the mission components could be permanently damaged or change their working properties. It is conceivable to have software which foresees all these situations, but the complexity of such a system would be too high. Therefore it is usual to wait until a failure or unexpected situation is discovered before updating the software.
- **Remote Diagnosis:** It is not always the case that the software can recognize and identify all possible anomalies. Hence, a remote diagnosis capability is required so that the developers can diagnose the system from the ground. The software must be able to collect state information on all subsystems and transmit it to the ground. This functionality is also used for large terrestrial machines (newspaper printers, excavators, mining machinery, packing machines, etc.). These machines are distributed worldwide, but maintenance engineers cannot be everywhere – they have to diagnose errors from remote offices.
- **High Computing Performance:** Usually the computing performance in spacecraft is slower than comparable devices on the ground by a factor of 10 to 100. The demanded high reliability and limited resources exclude high computing performance. Today, space applications are always in need of more computing power. New ways must be found to combine high computing performance with high reliability and limited resources (and not just for space applications).

These requirements are not new and they are also not so different from the ones valid for terrestrial applications. However, the problem with data processing systems in spacecraft is that all these requirements must be fulfilled at the same time.

4.6.1.2 Development Trends

Information technology (data management) is the fastest changing discipline for space missions. Hence, it is important to look not only at today's state of the art, but also to analyze the trends for future missions.

Trend 1: More Autonomy, More Data Reduction

In the simplest case it suffices if the data management collects, saves and later transmits the data to the Earth. However, in many cases we do not need the raw data, but only the **information** contained in it. For example, a satellite which should find a fire on the ground can take an optical image with 10 megapixels (30 Mbytes). However, we are only interested in knowing where the fire is located. This could be coded, say, in 1 kbyte. This is a data reduction of 30 000 to 1. Today all satellites transmit daily more than 10^{14} bytes to the ground. This is a huge amount, of data; however, of that, we need at most 10^8 bytes of information. Nevertheless, **data reduction** is very difficult and computationally intensive. Usually we receive almost raw data and extensive computations are necessary to extract the required information in computer centers on the ground, often with human help.

Here it becomes clear that information must have a purpose, in contrast to data. Only for a certain purpose or with a defined intention can one extract information from data. However, from the same data one can extract different information for different purposes.

The trend is to reduce the amounts of data to be stored and transmitted. This requires higher computing performance and smarter algorithms. The target is to save and transmit only the required information, and not the complete raw data. We can see this effect especially in hyperspectral and SAR (radar) images. These applications generate several gigabytes (10^9) or even terabytes (10^{12}) of data per second. To save and transmit the raw data would be a huge waste of resources. Hence the data processing and information extraction have to be done on-board of the spacecraft.

On the control system side the trend is to implement higher **autonomy** on-board the spacecraft so that operational expenses on the ground can be reduced. The spacecraft should be able to execute autonomously all actions to fulfill its mission over long time periods.

Today, some satellites can already plan their activities for the coming hours or days autonomously. This is of special importance for deep-space missions where human intervention is very difficult. An autonomous mission which is able to operate independently for several days must have high computing performance to be able to plan and reduce the accumulated data to (a few) required information bytes.

Ratio of Computing Performance to Frequency

Fortunately, the more complicated the operations are, the more time is available (not only needed) for the computations. For example, Figure 4.6.2 refers to the control of a robot arm. With a frequency of 5000 times per second sensors collect data like position and forces on the joints (5 kHz). These operations are very simple; it is just a matter of registering the signals and storing them.

The next layer with a frequency range around 1 kHz contains **steering** or **servo control** activities. The sensor data is filtered to reduce the noise, then a certain position, force or speed is held constant. Deviations must be compensated rapidly. Often the servo control is executed on a dedicated simple processing unit which is attached directly to the actuator (in this example, in every joint). These operations are more complicated than the ones described above, but they require a sampling frequency of only 1 kHz and not 5 kHz.

The next layer contains the **control activities**. The task is to determine the desired attitude, namely, desired position, desired force or desired speed, which should be maintained as predefined. These operations are more complicated; however, they may be executed with a lower frequency: between 100 and 2 times per second. In the example of the robot arm, this layer must coordinate all joints to reach a certain target or move in a desired trajectory. Another example is the attitude control of a satellite or spacecraft, where this layer must control, for example, three or four reaction wheels in such a way that a certain attitude is held or

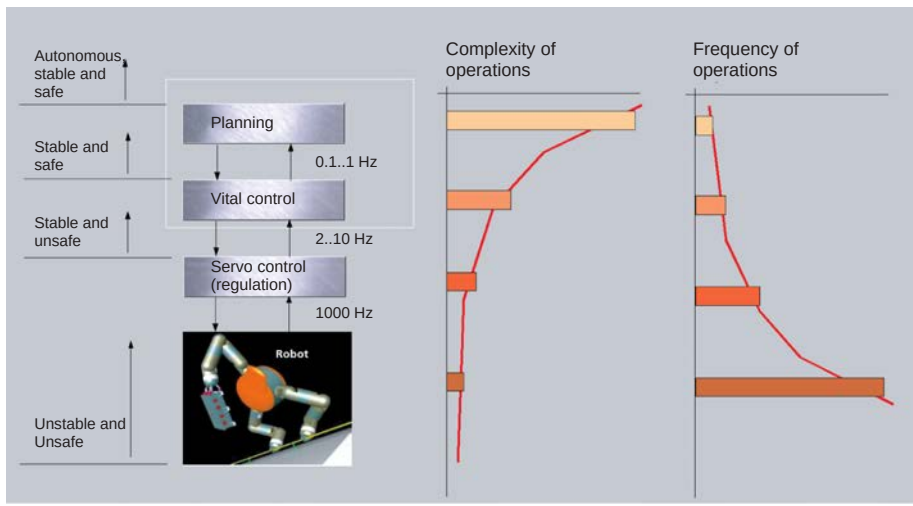


Figure 4.6.2: *Typical operations, their complexity and frequency.*

that the satellite turns to a certain position/attitude. The control must evaluate data from several sensors and thus control several actuators.

The next layer is the **planning** or **autonomy activities**. This activity comprises the planning of what activities must be performed during the next time period respecting predefined constraints; for example, the satellite should point its solar cells to the Sun, charge batteries and, when it flies over a certain goal, take and store images, then wait until it has a ground contact and transmit the images to the ground station. This could be a plan to be autonomously executed. Selecting the target coordinates for taking images could likewise be a job for the autonomy module. In this case the planning component must "know" the purpose of the mission and which constraints have to be kept. For example, the spacecraft should scan the complete Earth, excluding the oceans, but the battery charge may never fall below a certain value; in particular, the battery must be fully charged before entry into the shadow of the Earth. Today, operations on this layer are mostly executed by people; however, the trend is to execute the planning on-board automatically. The planning is a very complicated task, but it requires a lower frequency, in the range of minutes or hours.

Trend 2: Central versus Distributed Processing

Traditionally, many components produced by the spacecraft industry (likewise by car manufacturers) have their own control computer. As a consequence, the space segment contains many computers from many different manufacturers. The car industry has the same problem and tries to counteract this trend because the associated complexity is no longer controllable and the system becomes unreliable. There are many initiatives against this trend, and almost all automobile producers support one or more of them. As an example, **AUTOSAR** tries to create a uniform software environment. The aim is that the control of any device should be able to run and be executed on any computer in the car, and not just on a dedicated computer for the particular device. This trend already casts its shadow over the aerospace industry, in particular for micro-, pico- and nanosatellites, because the usual approach where "every device has its own control computer" can no longer be scaled down. Digital chips are already very small and economical, but this is still not sufficient for pico- and nanosatellites, where the available space, the allowed weight and the energy are very limited.

For very limited resources one must abandon the approach where "every device has its own controller," because it is not efficient. **Resource sharing** with intelligent software for **resource management** has to be implemented. The same hardware platform must be used for the control of many devices and for the execution of applications on the spacecraft bus as well as on the payload side. For example, on the main bus: telemetry, attitude control, navigation, power control, etc. For the payload we could have pattern recognition, image processing, classifications, etc. All these functions can be executed on one single computer.

For more complex spacecraft, different approaches are followed today. Some try to distribute the complete functionality to different processors, while others try to create a powerful central computer. Both approaches have advantages and disadvantages. With distributed functionality a failure will not immobilize the complete satellite, unless it concerns a vital function.

Another trend is to separate vital functions from extra (nice-to-have) functions. Vital functions (critical) are more difficult to be implemented and certified. Therefore, they should run totally independently of the nice-to-have functions. Preferably they should run on different processors to avoid any kind of interference. Another extreme is "a very critical function shall have its own processor." Some try to achieve the same goal by implementing several software virtual processors on a single processor (time-space partitioning).

With a **centralized** approach the on-board computer can be a **processor pool** which provides dependable storage and computing performance for all devices and applications. A network of (software) **tasks** can be distributed in this processor pool. Tasks replace traditional processors for each and every device. One could say that the tasks implement virtual processors for all devices. Payload tasks and other applications are likewise executed. This critical computer must be very dependable; **redundancy** features must exist. The computer pool must consist of several redundant independent computers. Even if some fail, those remaining should be able to fulfill the mission.

Trend 3: More Software

Not only in spacecraft, but in all technological areas there is a clear trend to move as much functionality as possible into software [4.6.1]. Figure 4.6.3 shows the rise and decline of control technologies, first implemented by means of mechanical parts, then going to more electronic (analog) components, and now toward software. The graphics refer only to the control and not to the function. It is not practical to replace the tires of a car with software; however, the complete motor control can be implemented with software.

Mechanical controls have been replaced by electronic controls, for example analog technology. Electronic controls are being replaced by software controls. The software runs on electronic components (computer elements like CPU, memory, etc.), but the functionality of such components is not increasing; they are limited to logical and arithmetic operations, storage and data transmission (networks).

Besides the traditional sequential software which is executed on a CPU, the next variant is emerging: **FPGA programming** (Free Programmable Gate Array programming). FPGAs are chips that provide millions of unconnected elementary logical operators (AND, OR, XOR, register). These operators are interconnected by programming. A complex functionality can be implemented by interconnecting such simple logical operators. The connections are configured not by hand but by a compiler for a special programming language, for example VHDL and Verilog. Then this program is translated (**synthesized**) to FPGA connections.

The shifting of functionality toward software is extremely interesting in space applications, because software has no mass, needs no space, uses no power and can be reloaded from the ground. But the introduction of software often presents more difficulties than expected. In many (small) space projects there are barely any software engineers; hence, the complexity of software development is underestimated, almost without exception. As a consequence the software is neglected at the beginning of project planning and toward the end the project is always on the critical path. This usually causes substantial schedule delays and higher costs.

4.6.1.3 Interconnection Topologies

In general the data management system consists of several communicating computers which are connected to several input and output devices. The interconnection structure is called **topology** and there are countless variations of such structures. Very simple devices without their own processor must be connected directly to a computer, perhaps using an analog/digital converter, or relays. More complex devices mostly have a (simple) integrated processor which implements a communication protocol. Communication with such devices is performed using

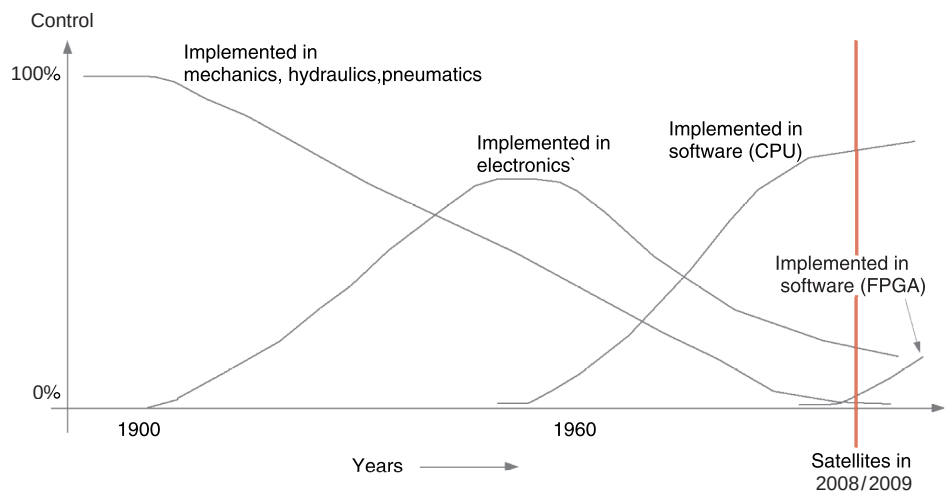


Figure 4.6.3: *Historical timeline of the application periods for different control technologies (not true scale).*

messages. Figure 4.6.4 shows some possibilities to connect devices to on-board computers.

With the **star topology** every device has its private connection to/from the on-board computer. This makes communication easier and safer, because the failure of a device has no effect on the communication with the other devices. The disadvantage is the large number of connections required.

A **bus topology** (e.g., **CAN bus**) is more economical, it needs only one connection to/from the on-board computer for many devices, and the number of lines is minimal. The disadvantage is that the failure of one device could block the entire communication exchange in the system. Hence, the bus has to be redundant. Furthermore, it should be possible to separate defective devices from the bus in case of failure. Another disadvantage is that all devices must have the same communication protocol (and the same speed). There are many standardized buses which are supported by many devices. Within embedded systems the CAN bus (originally from the automotive industries) is used very often.

However, the world is not black and white; mixtures of both topologies are mostly used.

For fault tolerance reasons the on-board computer consists mostly of several identical **computers communicating with each other**. The interconnection structure between communicating computers is based on a different principle than the interconnection between computers and devices. Here a master–slave hierarchy exists. The computer (**master**) controls and the device (**slave**) reacts accordingly. Within the interconnection structure for computers, all nodes (computers) are equal partners. Figure 4.6.5 shows some possibilities for interconnection topologies between on-board computers. Within star topologies and rings, redundant connections are often implemented as well.

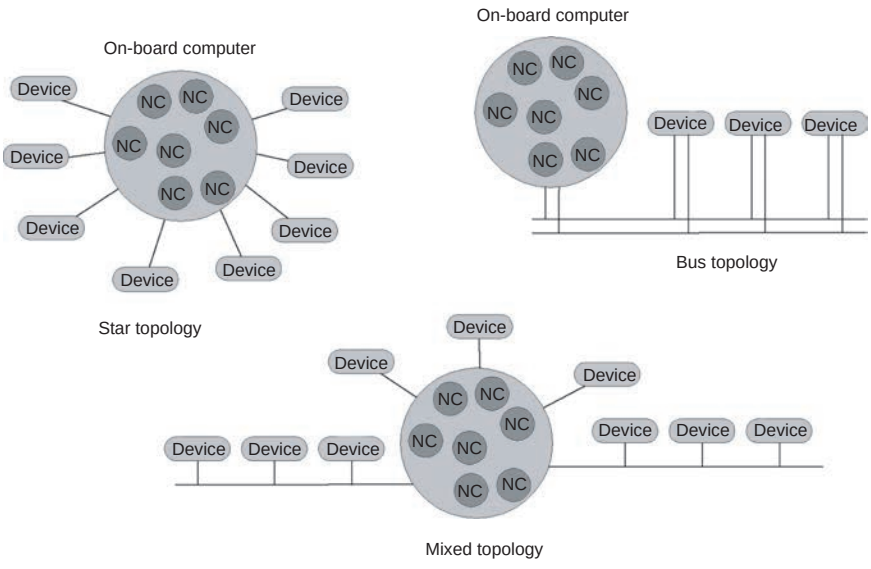


Figure 4.6.4: *Topologies to connect devices to an on-board computer (NC = Node Computer).*

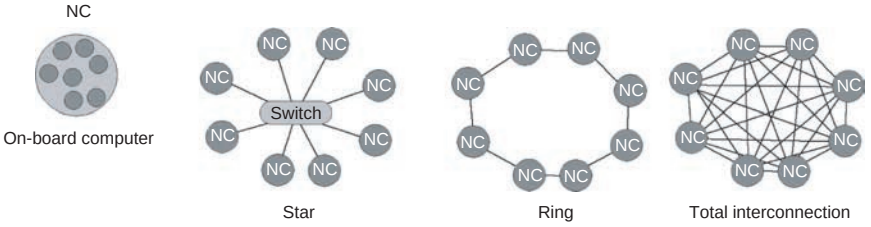


Figure 4.6.5: *Topologies for the interconnection of communicating computers (NC = Node Computer).*

4.6.2 On-board Computer

It is often postulated that the functions of the on-board computer are telecommanding, telemetry, attitude control, navigation, surveillance, payload interface, autonomy, etc. But if one takes a closer look, there is no module "surveillance" or "navigation" in the block diagram of the on-board computer. The hardware functionality is limited to logical/arithmetic operations, storage and data/signal transmission (networks), and these are precisely the modules one finds in the hardware design.

Although there are countless hardware architectures, a typical computer can be characterized as shown in Figure 4.6.6. This is a diagram of a single computer which is designated as part of a network of **communicating computers**. As said earlier, the on-board computer mostly consists of several such communicating computers.

4.6.2.1 Free Programmable Gate Arrays and System on One Chip

Recall that FPGAs are chips that provide millions of unconnected elementary logical operators (AND, OR, XOR, register). These operators are interconnected by programming. A complex function can be implemented by interconnecting such simple logical operators [4.6.2], [4.6.3], [4.6.4].

Especially in the area of input/output (I/O) and network interfaces, FPGAs are extremely interesting, because the variability of interfaces is huge and available components offer almost exactly all required interfaces. With FPGAs it is possible to freely define the function of every single pin and the internal functionality; it is also possible to program the complete I/O and network structure in one single chip, else one would need many chips and a large physical interface board (see Figure 4.6.7).

Besides the I/O functionality, it is common to shift filter and (pipeline) data processing into FPGAs. Operations which are executed repeatedly for a data stream are predestined for FPGA implementation.

Nowadays, FPGAs are so big that complete CPUs can be implemented inside them. Some manufacturers offer the possibility to synthesize different CPUs [4.6.3], [4.6.4], while others integrate up to four fixed CPUs into the FPGA chip [4.6.2]. It thereby becomes possible to think about **SOC** (**System on One Chip**). The target is to integrate the complete computer into a single chip. Figure 4.6.8 shows two approaches toward SOC.

FPGAs can implement almost the complete logic of the computer, but their storage possibilities are relatively limited. With an FPGA solution one must almost always attach external memory chips. Then one can no longer speak of a "system on one chip," but of a "system on very few chips." Alternatively, there are different manufacturers [4.6.8] which integrate CPU, flash (ROM), RAM, a network connection, different standardized I/O buses, and output signals in one single chip. They try to create a generic architecture usable for many different applications.

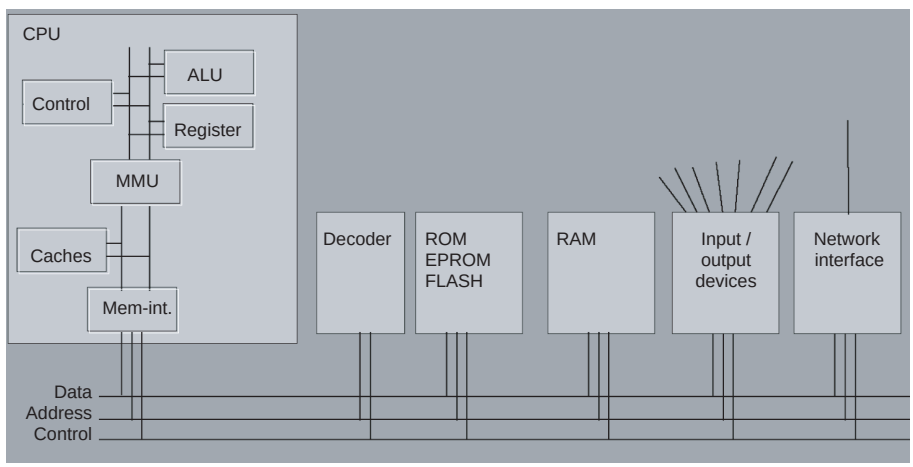


Figure 4.6.6: *Typical block diagram of a (node) computer. Abbreviations: CPU = Central Processor Unit; ALU = Arithmetic and Logic Unit; MMU = Memory Management Unit; Mem-Int = Memory Interface; ROM = Read-Only Memory; EPROM = Erasable Programmable Read-Only Memory; RAM = Random Access Memory (Read/Write).*

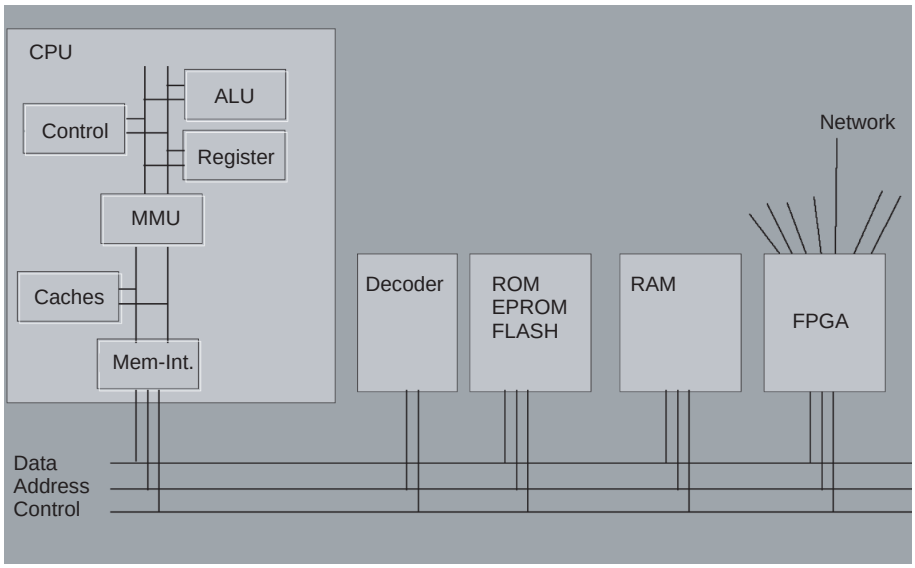


Figure 4.6.7: *Typical block diagram of a (node) computer in which the complete I/O structure is implemented in a FPGA.*

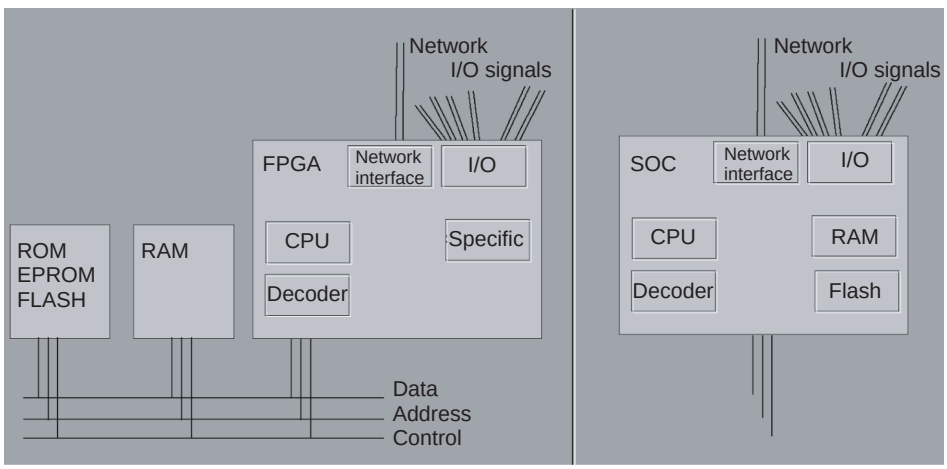


Figure 4.6.8: *Possible architectures for a "system on one chip" (SOC).*

4.6.2.2 Computers for Space Applications

So far we have made no difference between "normal" embedded systems and on-board computers for space missions. Extra precautions against radiation effects must be implemented for space applications. The easiest solution would be to block the radiation, for instance with a thick box of lead. However, this is not practicable because of the huge mass and the space requirements for the box. In addition, some radiation particles (**ions**) carrying very high energy could penetrate the box. Hence, other protective precautions are necessary.

Cosmic radiation consists of protons, alpha particles, heavy ions, electrons and neutrons. In miniaturized circuits three kinds of damage can be caused by these particles [4.6.9]: **totally ionizing dose** (TID) damage**, single event upset** (SEU) and **single event latchup** (SEL).

The TID causes rapid aging of the silicon components. The power consumption and the internal

noise therefore rise slowly but continuously, until the signal-to-noise ratio is so bad that the signal can no longer be clearly distinguished from the noise. If this occurs the component has failed permanently. The TID is measured (even today) in **krad** (kilorads of radiation absorbed dose, or the **energy dose**). The radiation intensity depends on the orbit. For low Earth orbits (below 800 km) one can assume 3 to 10 krad per year if protection is provided by a 2 mm aluminum shield. Normal electronic components survive approximately 30 krad and space-certified components (**radiation hardened**) are designed for a dose of 300 krad. For comparison, 1 krad is deadly for humans.

As mentioned, SEU and SEL are both **single event effects** (SEEs). They are caused by single high-energy particles impacting or penetrating the semiconductor material. The adjacent material absorbs the energy and an **ionization track** is created. If this happens within a flip-flop switch or a memory cell, a change of state might occur (from 1 to 0 or from 0 to 1). This causes a data corruption by SEU. SEUs cause only data damage; the component itself will not be damaged. One must expect 10 SEUs per day. To prevent this data damage, one has to implement some kind of redundancy code or create redundant data.

SELs are current spikes which can damage components permanently. A charged particle can cause an ion track in the substrate generating a very high current such that the chip surface is burnt. SELs are rare; several years can pass before one appears. To avoid damage, the circuit has to be interrupted immediately. In less than 1 ms the device (power) has to be switched off.

Figure 4.6.9 gives a summary of measures for designing safe space computers. The only possible precautions against TID are to screen the parts used and to select robust hardware components. Data redundancy helps in recognizing corrupted data and correcting it as a measure to counteract SEUs. As protection against SEL, the complete circuit must be switched off within a millisecond after detection of the occurrence. A couple of seconds later it can be switched on again.

SEL can cause a sudden power spike which under certain circumstances might be only a little higher than the normal current. The **latchup protection** (**current monitoring**) must recognize such spikes and must switch off everything immediately. To be able to recognize the spike it is advisable to monitor all components separately.

Data in memory and in FPGA registers can be protected by redundancy. Redundant codes (EDC codes) allow data corruption to be recognized and 1 bit per entry (register) to be corrected. The simplest way to verify the correct execution of the logic is to replicate it two or three times on-board and compare all results.

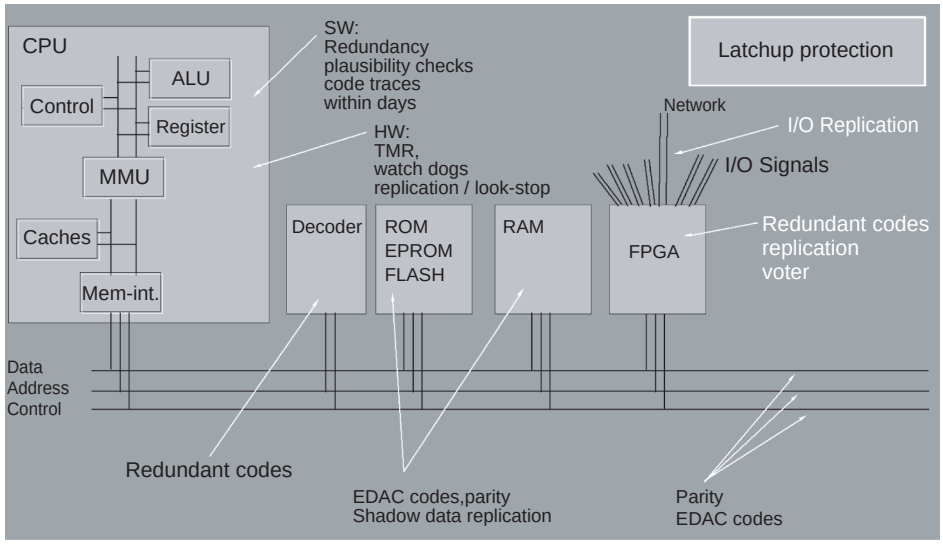


Figure 4.6.9: *Measures to protect a computer against the effects of cosmic radiation.*

With standard CPUs one cannot expect such protective measures, hence ESA has developed the LEON-FT processor [4.6.12]. Inside LEON-FT every function and every register is replicated three times (**TMR**, **Triple-Module Redundancy**). After every operation all three results are compared; if one is different, it is immediately corrected. LEON-FT is safe, but slow. Another alternative is to use standard **COTS** (**Commercial Off The Shelf**) processors and to build the required redundancy externally or by replication in software. For example, two or three processors can gather their results and compare them. With the software one can execute every function two or three times and compare the results. It is more economical to check only results against plausibility rules. A very simple and economical measure is to use **watchdogs**; for every function a maximum allowable duration is defined. The watchdog has to be triggered within this time period, otherwise it will execute a reset of the computer. After this a recovery action has to take place.

4.6.3 Software

Software is only information and has no physical representation or properties. Because the subsystems of a spacecraft are measured according to their physical properties (mass, space, power consumption, life expectations, imaging, etc.), it is very difficult to treat software as a subsystem like all the others. In addition, there are no unequivocal metrics for software. Today there are approx. 400 **metrics** to measure the properties of software, but they are very controversial and are not generally accepted.

All this makes software rather unfathomable. One only knows that software is immaterial, abstract, complex, everywhere, big, expensive and error prone. Dealing with mechanical components, one can recognize complexity and thorough craftwork at first sight. With software it is not so simple. Nevertheless, software is at the core of every space mission and the fastest growing component measured in terms of complexity and development cost. The same applies to the car industry. While 22% of the added value of a car in 2004 was contributed by software, this value is expected to increase to 36% in 2009.

4.6.3.1 Problems with Software

Progress in information technology is rapid: chip capacity doubles every 18 months and the systems can be made smaller. The communication bandwidth doubles every 12 months; the systems can thus become quicker and quicker. Software complexity and size double every 10 months. Software grows faster than the hardware capacity. It is predicted that the time will come when systems will grow in physical size and become slower instead of faster (see Figure 4.6.10).

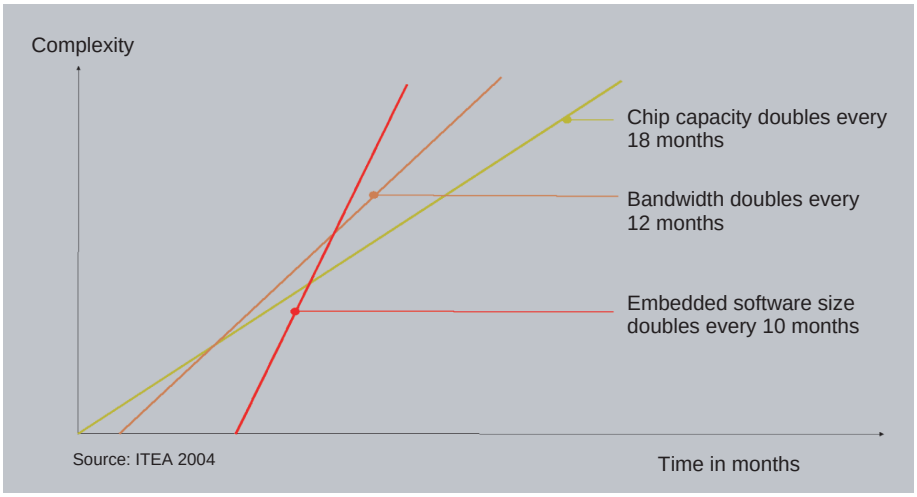


Figure 4.6.10: *Growth of chip capacity, communication bandwidth and software complexity (logarithmic scale).*

Table 4.6.1: *Software complexity of selected systems in lines of code (LOC).*

System	Software Complexity
BIRD satellite	90 K-LOC
TerraSAR satellite	90 K-LOC
Control of A310	400 K-LOC
Cellular phone	1 M-LOC
Cars	2 M-LOC
Control of A340	20 M-LOC
SAP	30 M-LOC
Linux	30 M-LOC
Windows	45 M-LOC

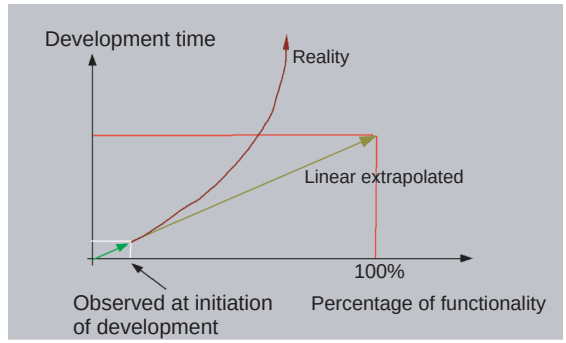


Figure 4.6.11: *Typical misjudgment of software development times.*

4

Software complexity grows much faster than its functionality. For example, Linux grew within three years from 15 million lines of code (**LOC**) to 30 million LOC (M-LOC). Windows 2000 has grown from 20 M-LOC to 45 M-LOC for XP (the size of Vista is a secret). However, the functionality has not doubled, only the complexity. Until this day there is no measure for complexity, hence lines of code are used. For comparison, the human capacity to completely understand a program code probably lies at 5000 **LOC** (5 K-LOC) [4.6.5] (see Table 4.6.1).

It becomes clear that no one can handle this complexity completely; hence, it is currently not possible to develop perfect software. The biggest challenge in software development is to reduce and control the complexity.

Besides complexity, the next big problem with software development is the required **know-how of the developers**. Spacecraft software contains the total knowledge of the complete mission. This is the combined knowledge of the involved computer scientists, electrical engineers, mathematicians, physicists, mechanical engineers and space engineers. This interdisciplinary collaboration often introduces more difficulties than originally expected, because all these specialists speak different (technical) languages and it takes a long time to understand all the technical aspects.

Software is decisive for the success or failure of a space mission. Many mission failures can be traced back to software errors. Software is usually the contribution finished last, and the costs for software development are often twice as high as expected or even higher. The reason is not that software developers are especially negligent, but that:

- Software complexity outstrips human mental capacity.
- Software is still neglected in project planning and is totally underestimated.
- The available time for software testing is usually too short.

Normally the testing of critical software consumes up to 80% of the software development costs. The testing of software functionality is the most difficult and time-consuming test for every space mission, because it contains the complete spacecraft behavior. The solution is not to avoid software implementation, but to reduce the complexity, and not to underestimate the complexity. Figure 4.6.11 shows the typical misjudgment of software development times.

4.6.3.2 An Example for Reducing Software Complexity

Modern aircraft and rockets no longer have physical or mechanical stability. Without software support it is no longer possible to fly them. The complex functionality of modern satellites cannot be implemented without software. As mentioned previously, the solution is not to do without software, but rather to reduce its complexity and not to underestimate it. But reducing complexity is in direct conflict with

the increasing demands being placed on that very software. The example of the BIRD satellite will be used to show how complexity can nevertheless be reduced.

With the goal of reducing complexity, the BIRD Operating System (BOSS) [4.6.6] was originally developed for the BIRD mission. BOSS and its **middleware** (see Section 4.6.3.4) were developed with the goals of achieving the highest possible reliability and implementing a level of complexity which cannot be reduced further. For every function there is a theoretical lower limit of complexity required for its realization. It cannot be made simpler without destroying the functionality, but it can always be made more complex. Finding the sweet spot of no longer reducible complexity is extraordinarily difficult, but remains the goal of BOSS.

BOSS consists of about 4000 lines of code. Current real-time kernels consist of about 1 million lines of code. Nevertheless, BOSS and its middleware have all the functionality required to implement embedded real-time systems. In contrast to the current trend, BOSS and its middleware are getting smaller without losing functionality. The current version of BOSS has about half the lines of code of the first version, although the current functionality is higher than before. For every single component the question is asked whether it can be made simpler, while nobody is trying to make it more complex (or elegant). This procedure goes against the current trend. The important point is that every developer can understand the entire BOSS implementation, and does not have to assume anything as a black box.

4.6.3.3 Software Functions and Requirements

Payload or mission science software is somehow different from the bus control software which runs on the on-board computer. The function of the payload software is **data reduction** and data management. The payload software is supposed to extract information from the data provided by the sensors (e.g., cameras, radar and other sensors). These applications are mostly numerical computations or number-crunching tasks with a large amount of data. Some applications require **on-the-fly** data processing. This means that sensor data should be processed as soon as it is provided by the sensor, that is before the data is stored. This is the case when the data rate is too high for storage. In other cases it is possible to first store data from the sensors and then to process it in batches, maybe in the spacecraft or maybe on the ground.

The satellite bus control software (in the on-board computer) usually fulfills mostly control and steering functions, which should be computed on a real-time basis. Typically these functions are split into three groups:

1. Booting, initialization and loading programs
2. Real-time operating system
3. Applications.

The **boot program** (**booter**) must reside in a non-volatile memory (e.g., flash or ROM). After every power-on and every reset command the boot program is activated and executed before any other function. It must initialize the hardware and then load the operating system and, if foreseen, the appropriate applications. When loading new software, the booter copies it to memory (RAM), initializes it and jumps to the start entry of the loaded software. Then the loaded software takes control. The operating system can likewise be saved in the same flash memory, or the boot program can load it from the network or disk. For space applications the boot program and the operations system usually are stored on the same flash memory.

The **real-time operating system** administers the resources of the computer, the time and the applications. It controls the I/O devices using I/O drivers. Modern operating systems provide **middleware** which implements the communications between the applications. Middleware helps make it possible for applications to be developed and operated independently of each other. Without middleware the applications would have to communicate directly with each other, often by means of common variables. This method results in a strong dependence between applications, which makes the system more complex.

The (user) **applications** run on top of the operating system and middleware. They have the task to check and control the different subsystems of the spacecraft. Applications for the payload segment vary for every mission, as explained above. However, for

spacecraft control the following standard software tasks can be identified:

1. **Initialization after Separation from the Launcher:** For example, check the subsystems, determine the orbit and deploy the solar panels after the spacecraft is stabilized. Nonvital devices (e.g. payloads) stay switched off. They are activated and tested later under the control of the ground station personnel.
2. **Housekeeping:** Collect and store status information from all subsystems, plus other software tasks.
3. **Telemetry:** During each pass transmit the current state of the spacecraft, the housekeeping data and payload data to the ground station.
4. **Command Interpreter:** Receive commands from the ground station, check, interpret and forward them at the right time to the corresponding subsystem or application for execution.
5. **Time Manager:** Many activities in the spacecraft are controlled by a timeline. The time manager is responsible for distributing a correct and consistent time throughout the entire system. The time tags from GPS messages are often used to synchronize the on-board time.
6. **Supervision Module (Surveillance):** The surveillance module checks the current state of all vital subsystems and compares the relevant parameters with predefined limits which should not be exceeded during nominal operations. If a critical deviation is detected, an emergency procedure is initiated. The emergency procedure often consists of the processing of a predefined command list (**safe list**). This list contains commands to switch off all unnecessary devices and to guarantee with the highest priority the survival of the vehicle.
7. **Attitude Control:** Orientation of the spacecraft (see Section 4.5.).
8. **Navigation:** Determination of the orbital position and the trajectory (see Section 2.2.).
9. **Power Management:** Checking the battery charge state and the energy balance of the complete vehicle (see Section 4.2.).
10. **Temperature Control:** Taking care of the temperatures within the vehicle to avoid supercooling or overheating (see Section 4.3.).
11. **Guidance Control:** Responsibility for the control of orbit maneuvers (see Section 3.3.).
12. **Software Upload:** Usually during a mission the software or parts of it must be replaced by new versions. The software updater accepts new software versions (uploads) from the ground, checks them and replaces the old ones.
13. **Autonomy:** The autonomy module "knows" the purpose of the mission and tries to fulfill this purpose with minimum human intervention. The autonomy module provides mission operations automation.

4.6.3.4 Real-Time Programming

Real time does not necessarily signify a quick reaction, but it means to guarantee a reaction inside a certain time window [4.6.7]. The correctness of **real-time tasks** not only consists of the generation of a correct output value, but also has to be generated at the right point in time and not simply "as fast as possible," such as an alarm clock which should ring around 06:15: its correct output is not only "riiinnnnggg," but "riinnnggg" around 06:15, it must not be exactly 06:15:00.000; there is a **response time window** with a certain tolerance, for example the ring would be satisfactory between 6:12 and 6:18. The end of the time window is the deadline (not later). In connection with real-time definitions, only the **deadlines** are mentioned; the beginning of the time windows is often ignored. For example, if somebody wants to jump onto a train, it is bad to jump too late (behind the train), but even worse to jump too early (in front of the train)!

"Real time" says nothing about speed. Some applications have very long **periods** and other very short periods. For example, an alarm clock which must ring at a certain time every day would be a real-time application with a period of 24 hours. A charging system which should print invoices at the end every month and mail them is likewise a real-time application with a frequency of 12 times per year. In contrast, the frequency control of an electron accelerator can be up to 1 million times per second (1 MHz). Typical control tasks in satellites have frequencies in the range of 1 Hz (not very fast).

In general computer-controlled tasks are implemented as parallel **tasks reacting in real time**. Some are **time triggered** as explained before, and others

react to external events (**reactive tasks**). The execution of reactive tasks can be split into three phases which are activated by an event or time point:

1. Read sensor measurements: the current state of the controlled equipment is determined.
2. Compute: the control reaction to the events is computed.
3. Address the actuators: the computed reaction is implemented as an action.

The periods and the limitations of the time requirements are different for steering tasks and for control tasks (see Figure 4.6.2). The steering tasks have a relatively high frequency ranging from 100 Hz up to 30 kHz (30 000 times per second) but are easy to calculate. In space applications a frequency of 1 kHz (within the servo control of the devices) is very common. The control tasks have relatively low frequencies, but are more complicated to calculate. Their frequency might range between 1 Hz and 1000 Hz. In space applications a frequency of 2 Hz is usual. Thermal checks, for example, can have in contrast a period of 30 seconds (temperature changes relatively slowly).

Classification of Real-Time Tasks

Typically a control job is split into tasks running in parallel. Some of them can run **periodically** (**cyclic**) and others **aperiodically**. For periodic tasks the three phases mentioned above have to fit into a given time window.

The aperiodic tasks are activated as **sporadic** tasks by asynchronous **external events** (**interrupts**) or as **spontaneous** reactions to **internal events**. All tasks have an initialization function which is called up once at system start, and a "handle" function which is called up for every activation of the task. The **ongoing** tasks run in the background without time requirements. They execute long-time computations which are not time constrained or check system states whenever the computer is in idle mode.

Time-critical or **hard real-time tasks** have fixed deadlines or response time windows which may not be missed. If the reaction does not take place in the intended response time windows, the system fails, causing unwanted effects: for example, not switching off the heat source of a boiler before it explodes. Switching off a millisecond after this hard deadline is of no help.

For the **soft real-time tasks** or **time-sensitive tasks** the response time window is not defined as strictly as for the time-critical tasks. The response time window is to be interpreted rather as a desirable time. For example, turning on a heating system if the temperature goes below 19 °C is not so time critical as the case of avoiding an explosion. If it takes place some seconds or even minutes later or earlier, it does not mean that the system has failed.

4.6.4 Dependability

Spacecraft anomalies and failures have to be expected. However, the spacecraft must continue its mission despite anomalies. This is called dependability [4.6.7]. Dependability, as shown in Figure 4.6.12, consists of:

- **Safety:** No catastrophic consequences and no danger should come from the system.
- **Reliability:** Continuity of service without interruption.
- **Availability:** Readiness of the system to be used.
- **Confidentiality:** Protection against information abuse.
- **Integrity:** Protection against information corruption.

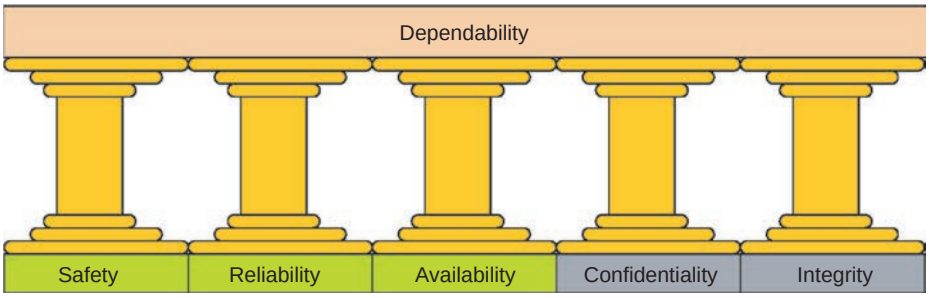


Figure 4.6.12: *Pillars of dependability.*

When a failure of the control system can have catastrophic or life-threatening consequences the control system has to provide **ultrahigh dependability**: This means a maximum of 10^{-9} failures per hour may occur, or, in other words, a maximum of one failure in 10^9 hours of operation is allowed. This means only one failure in 114 000 years of operation! Such ultrahigh dependability cannot be guaranteed with today's technology, but many control systems have to be (and are) certified according to this measure. In the satellite business a failure has (normally) no life-threatening consequences, only loss of high investments. For such systems a dependability of "only" 10^{-6} to 10^{-7} failures per hour is tolerable. In contrast, human space missions require a dependability of 10^{-9} failures per hour.

To achieve the highest possible dependability in embedded control systems and to avoid possible failures it is recommended to:

1. Minimize the probability of malformations.
2. Recognize and handle anomalies while the system is running.

To satisfy these two points the three approaches listed below should be followed [4.6.6]:

1. Search for the lowest possible complexity: make the design as simple as possible because the main cause of malfunctions and failures is high complexity.
2. Implement fast and safe recovery strategies for each possible anomaly.
3. Optimize redundancy and redundancy management. It is possible to have reliable systems in spite of unreliable components. Redundancy alone does not enhance dependability; more important is redundancy management, which decisively influences the design of software and hardware from the beginning to the end of the development.

If a system has been certified as very reliable and safe it means only that very few failures can occur. Absolute reliability (a system which never fails) is not possible in technical systems. Always be aware of anomalies!

An **error** is the cause of undesirable results and can lie with people (operators) or machines. The environment can also introduce unexpected circumstances causing undesirable results. All this is summarized in the term **anomalies** (errors and surprises).

All components of a system are error prone. But not all components are endangered in the same way. Mechanical errors which can lead to catastrophic consequences are primarily caused by control errors. Most of them come from unpredictable situations. A second category of failures is caused by mistakes in the development (mostly in the software, because of its complexity). A third source for malfunctions is mechanical wear and tear as well as failures of electronic components.

Errors can be classified as follows:

- Source: development errors or runtime error
- Behavior: permanent or sporadic errors
- Category: value or time errors.

A **fault-tolerant system** should fulfill its function even after the occurrence of faults, errors and internal failures. It may contain errors; however, it may not propagate them or make them visible. Any internal error, failure or disturbances must not show external effects.

A **safe system** may interrupt/discontinue its function, but it may not pose any danger. If a fault-tolerant system can no longer perform its function, say, after many failures, control of the system cannot simply be lost; the system must rather be configured to a safe, stable state and remain there until help can be provided.

Active modules, such as processors, are masters and can act independently. In contrast, slave modules (simple I/O devices, sensors, actuators) do not act independently; they only react to commands from master modules. Therefore, anomalies from active modules can spread faster in the system and cause further anomalies in other modules. An active module is able to disturb other modules by sending messages, bus accesses, entries in other memories, etc. The consequences of these anomalies are more dramatic when the modules also directly control the actuators.

Module errors can be handled internally in each module; otherwise, errors will be visible outside the module (they have visible consequences) and must be handled externally. For internal error handling the processors and all other module components must implement redundancy requirements. For external error handling the modules can be implemented to be redundant by replication. The results and

actions of redundant modules must be compared permanently to detect and handle anomalies. This comparison function (**voter**) can be implemented as dedicated hardware or executed in the same processor modules.

4.6.4.1 Module-Internal Handling of Errors

Internal error detection and handling must cover all module components:

1. The processor, mostly in the form of a microcontroller
2. The memory
3. I/O and network connections.

Normal microcontrollers have a high degree of integration, but normally no support for internal error detection and handling. Coarse mistakes, for example if the processor gets stuck or goes astray, can be detected by a **watchdog logic**. A counteraction could be to perform a module reset. For a more sophisticated supervision scheme one must duplicate (or even triplicate) the central processing unit (duplex central processing unit). In this case all replicas must perform the same function at the same time; therefore, the behavior of the CPUs can be compared externally.

Transient internal CPU errors can be handled by executing every software task twice. If both executions produce the same results, it can be assumed that there was no error. If they differ, the same task can be executed a third time and compared again to identify the (probably) right result.

LEON-FT from the ESA CPU was developed especially for safety-critical systems. In LEON-FT every internal function is implemented three times (triplication) and the results are compared immediately after each basic operation. Deviations are corrected immediately.

Memory and storage errors are simpler to detect and correct. For securing memory and storage devices we can use simple **parity**, **EDAC/EDC codes** (Error Detection and Correction codes), **CRC** (Cyclic Redundant Code) at block level, multiple copies of the data, and address range protection registers (rather for software errors).

Bit flips in memories accumulate with time (new flips can occur at any time). In general EDAC can only correct one bit flip per word. If, however, two bit flips have accumulated in a single word, we can still detect the error, but (normally) we cannot correct it. Therefore, the complete memory should be examined cyclically and every bit flip corrected as soon as possible, as long as it is possible.

4.6.4.2 Module-External Handling by Using Redundant Modules

The most common and simplest method for external error handling is to replicate the modules as a whole and externally compare their results by using a supervisor [4.6.7]. The supervisor needs a reference to be able to recognize anomalies, for example for plausibility checks or for simply comparing the results of several redundant modules. But note especially that the problem of **comparison** is difficult, because small input differences can lead to absolutely different results. This happens with values near **decision thresholds**, for example

```
heater() {temperature = readTemperature();
        if (temperature > 19.5) { ... } else { ... }}
```

With a value around 19.5 a sensor tolerance of only 0.01% can lead to huge differences, so that the results of two parallel running modules would no longer be comparable. To solve this problem, the **voter/comparator** needs more information: for example, whether a decision was near a threshold value. This makes the comparison of results much more difficult.

The voter can be implemented as a dedicated passive module (hardware) observing the active modules. This increases the complexity of the system, which brings new risks. What will detect a malfunction of this new module (voter)? However, the voter module is much simpler than the active modules and therefore fewer errors can be expected there. In any case the voter can become a **single point of global failure**.

The comparison task can also be performed by the active modules themselves by monitoring each other and coming to a democratic (voting) decision if there are differences. This method needs a longer decision time than a dedicated voter, but the implementation of a more complex comparing mechanism is possible.

An external supervisor recognizes errors, indeed, later than a module-internal supervisor; however, it has more decision possibilities, such as for example

the comparison of several modules, timing checks, synchronization points, communication protocols and checks of transmitted messages, and it also has the possibility to neutralize uncooperative or suspicious modules (e.g., by resetting them). Figure 4.6.13 shows two possibilities to isolate uncooperative modules:

1. Each bus connection of the modules has a tristate buffer which is under the control of the supervision unit.
2. All bus connections to the bus go through resistors (insulators), and the supervision unit is able to interrupt the power of every module separately. This method is more economic and safer, and one can save power. But the bus becomes slower and turning off modules implies loss of information and requires a more complex recovery mechanism.

4.6.4.3 Self-checking Pairs

By using only two identical, redundant, active modules errors can be detected by comparing results, but they cannot be corrected. We see the difference, but we cannot say which one is right and which is wrong. By using three redundant modules errors can be detected, localized and corrected (two of three, a democratic decision), so that a single error has no effect on the following superior modules [4.6.7]. By using two redundant modules, a **self-checking pair** or **trustworthy module** can be built up. A trustworthy module is partially correct so that it supplies only correct results, or none in the case of error. A trustworthy module alone is not fault tolerant, but it is a good component to build fault tolerance. It can also be used only as a **fail-safe** module. With this configuration the module should supply a safe reaction (e.g., emergency stop) or start an emergency program if an error is detected.

Figure 4.6.14 shows three possibilities for trustworthy modules (A, B, C). Each module supplies not only a value (result), but also a signal (OK) to confirm the correctness of the value. This signal can also be used to isolate modules from the rest of the system in case of (internal) errors.

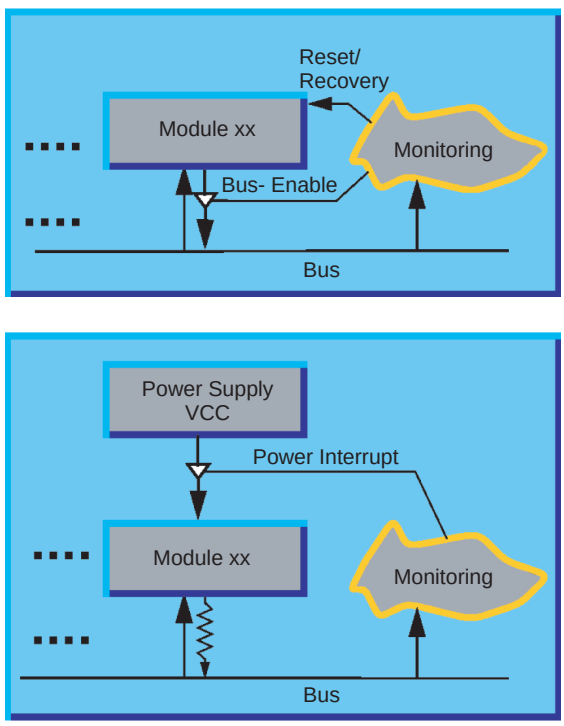


Figure 4.6.13: *Circuits for isolating faulty modules.*

4.6.4.4 Fault-Tolerant Structures

A fault-tolerant module can be built by using of two trustworthy modules [4.6.7]. One of them is the **primary** module (**worker**), the other one is a **shadow** module (**monitor**). The monitor runs in parallel with the worker, but does not interfere. The primary module has control of the system until it recognizes an internal error or until it fails. In this case the failed module keeps silent and executes a recovery. The shadow module detects the failure of the worker and immediately takes over the role of the worker, assuming the control function for the system. When the first module is again ready for operation, it takes the roll of the shadow (monitor) module and remains silent in readiness. If the shadow module recognizes an internal error, it executes a recovery in silence. The shadow module becomes active when it notices that the expected reaction ("OK" signal) of the primary module does not arrive on time or after an agreement between both modules (communication protocols). A similar approach was implemented in the BIRD satellite. In BIRD every module checks itself by software so that the complete construction requires only

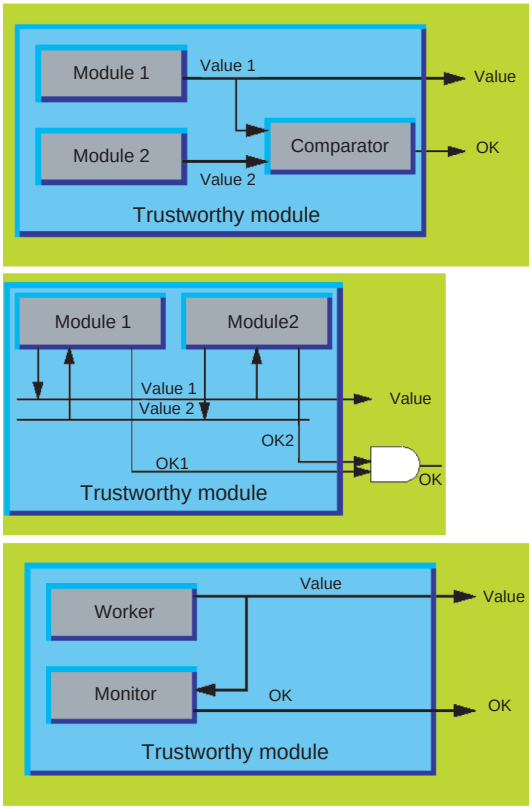


Figure 4.6.14: *Three methods for constructing trustworthy modules.*

two instead of four computers. The error coverage is lower, but for a satellite mission it is high enough [4.6.10], [4.6.11].

It becomes unpleasant if both trustworthy modules find an internal error at the same time. Then both withdraw to carry out a recovery and the system remains without control. This can happen, for example, if both trustworthy modules are identical and have the same development error. Ironically it would have been better not to recognize the error because to do nothing (no reaction of the control) can be worse than doing something wrong.

It is also possible to construct directly fault-tolerant modules without using trustworthy modules. This is based on democratic decisions with three or more redundant modules. The basis configuration consists of **triple modular redundancy** (**TMR**), as shown in Figure 4.6.15.

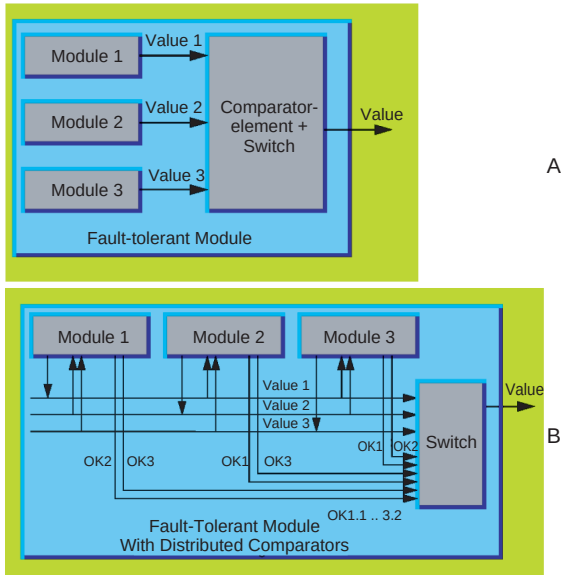


Figure 4.6.15: *Configurations for triple modular redundancy.*

With a TMR configuration three similar (identical) modules are used. They execute the same task. A voter checks whether all three modules provide identical results. If this is the case, the result is forwarded. If one result differs, it will be masked out (ignored) and the value of the two identical results will be forwarded (democratically). If all three results are different, then we have a big problem. But beware: the fact that two results are identical does not necessarily mean that they are correct, only that they are very probably correct. The advantage of the TMR technology is that (apparently) faulty results can be masked "on the fly" to keep the timing without time delay.

Bibliography

[4.6.1] ITEA. European Leadership in Software-intensive Systems. http://www.itea-office.org/, 2003, 2004, 2005, 2006.

[4.6.2] Xilinx. Products and Company Profile. www.xilinx.com, 2006.

[4.6.3] Altera. Products and Company Profile. www.altera.com, 2006.

[4.6.4] Alcatel. Products and Company profile. www.alcatel.com, 2006.

[4.6.5] Montenegro, S. Software Komplexität. Fraunhofer IESE: Online-Forum VSEK, Software Engineering, www.vsek.org, 2007.
[4.6.6] Montenegro, S. BOSS-Documentation. Property: FhG FIRST. To be delivered upon request (info@first.fhg.de), 1999, 2000, 2003, 2004, 2006.
[4.6.7] Montenegro, S. *Sichere und fehlertolerante Steuerungen.* Munich: Carl Hanser Verlag, 1999.
[4.6.8] Barth, J. The Radiation Environment. http://radhome.gsfc.nasa.gov/radhome/environ.htm, 1998.
[4.6.9] ARM. LPC2000. http://www.mikrocontroller.net/articles/LPC2000_Philips_ARM7TDMI-Familie, 2008.
[4.6.10] Briess, K., Baerwald, W., Gill, E. *et al.* Technology Demonstration by the BIRD-mission. 4th IAA Symposium on Small Satellites for Earth Observation, April 7–11, 2003.
[4.6.11] Baerwald, W., Montenegro, S. PowerBird – Modern Spacecraft Bus Controller. *Acta Astronautica*, **52**, 957–963, 2003.
[4.6.12] Leon-Processor Dokumentation. http://www.iti.uni-stuttgart.de/~virazela/LP_Project/leon-2.3.7.pdf, 2004.

4.7 Communication

Peter Turner

An extract from the large area of communication technologies for space systems is presented. Although the following considerations refer to satellite communication, they are in principle valid for many other applications.

For this reason it is necessary to address some fundamental principles of telecommunication to facilitate the understanding of interrelations and physical boundary conditions. At this point consideration of theoretical channel capacity and the influences of omnipresent thermal noise are of elementary importance.

The specifications of the transmission system with respect to capacity, reliability and compatibility with the ground stations are crucial for the architecture of the on-board system. They determine the size and characteristics of the on-board antennas, the necessary transmitting power and the sensitivity of the receivers. In addition, coding and modulation are directly connected to the achievable transmission rate, quality and data integrity.

Finally, the fully redundant communication system of the BIRD satellite is discussed, which has been successfully orbiting Earth since 2001.

All subjects addressed in the following sections can only provide a small insight into the large area of telecommunication.

4.7.1 Introduction

Nowadays satellites are a central part of daily communication. Intercontinental telephone and data links as well as television transmissions are constantly in use and subject to competition[1] amongst various providers using various technologies. Hence the **efficiency of the communication system** is a significant factor

[1] Especially by intercontinental submarine cable networks, employing modern glass fiber and modulation technologies with transmission bandwidths in excess of 10 terabyte per second.

for its design and implementation. As many users as possible should be able to use the same satellite link simultaneously. Therefore, one tries to place as many channels into the allocated frequency band as possible. With the help of modern modulation methods one comes constantly closer to meeting these requirements. Limiting factors are the available primary power (e.g., 7 kW for ASTRA 2A) of the satellite bus, the associated thermal budget and the various regulations regarding the radiated radio frequency.

In contrast to digital data links, the delay of voice transmissions via geostationary satellite links has a perturbing effect, especially when an additional **echo** is present. Modern technology is able to cancel a large part of this echo but reaches its limits when links extend over two satellites with a corresponding doubling of the delay.

Modern telecommunications satellites have a **design lifetime** of up to 15 years but in reality their demise is determined by the depletion of the propellant available for attitude and position control. In general, the lifetime of modern electronic components is longer in most cases, which has been demonstrated by elaborate and costly tests. In addition vital subsystems (e.g., transmitter, receiver and transponder) are designed with full redundancy that can be activated when needed.

The technology installed on-board a satellite must be compatible with the transmit and receive systems of the allocated ground stations, which might consist of a large number of single stations distributed worldwide.

For short-life (< 5 years) research satellites, considerably lower cost components (commercial off the shelf (COTS)) and systems from other areas of space technology are being used at an increasing rate.

At the end of this chapter the transmit and receive system of the DLR BIRD (Bispectral Infra-Red Detection) research satellite is presented, whose technology is mainly derived from the area of sounding rockets. This system has been working faultlessly for more than seven years in orbit.

4.7.2 Radio Spectrum

In addition to national institutions, the European Cooperation for Space Standardization (ECSS) determines all frequencies for communication applications in Europe. The International Telecommunication Union, Geneva (ITU), is responsible for the allocation and coordination of frequencies (Table 4.7.1). This results in partially overlapping and controversial band allocations for the assigned sectors. Furthermore, one has to consider that special frequency allocations exist for various world zones. Many frequency allocations for satellite communication are also used for terrestrial applications. Due to the permanently growing number of services and to disturbances by "human-made noise," there is a trend to switch to higher frequencies. **Detailed frequency allocations** are given in the publications of the ECSS and ITU (Article 8, Radio Regulations). Another breakdown can also be found in Section 6.3.

Table 4.7.1: *Excerpt from a list of allocated frequencies for various satellite services.*

Band	Frequency [MHz]	Service	Direction
S	2025–2120	SR, SO, EES	Earth – Space
	2200–2300	SR, SO, EES	Space – Earth
X	7145–7235	SR, SO	Earth – Space
	8025–8400	EES	Space – Earth
	8400–8500	SR	Space – Earth
K_u	10.7–11.7	FSS	
	12.5–12.75	BSS	Space – Earth
	11.7–12.2	SMS	Earth – Space
	14 000–15 350	SR	
	16 600–17 100	SR	
K_a	25 500–27 000	EES	Space – Earth
	31 800–32 300	SR	Space – Earth
	34 200–34 700	SR	Earth – Space
	37 000–38 000	SR	Space – Earth
	40 000–40 500	SR, EES	Earth – Space

4.7.2.1 Atmospheric Attenuation

Higher frequencies (= shorter wavelengths) enable the use of relatively small parabolic antennas (e.g., for the ASTRA TV satellites), as the **physical size** of an antenna (effective area) is directly proportional to the wavelength used. However, one has to consider that **shadowing effects caused by obstructions** (no direct line of sight to the satellite) increasingly play a role when high frequencies ($f > 1$ GHz) are used and in extreme cases can result in a total loss of

communication. The economic use of very high frequencies ($f > 20$ GHz) is restricted by the atmospheric **absorption bands** caused by water vapor and oxygen, which leads to high link attenuation.

4.7.2.2 Maximum Power Flux Density (from Space to Earth)

Every signal transmitted by a satellite is governed by the rules established by the ECSS. Besides the potential modulation and transmitting bandwidth, the **equivalent isotropic radiated power** (EIRP) has to be limited to comply with the maximum permitted incident power flux density produced at the Earth's surface. This power flux density depends on the elevation of the satellite and must stay below a predetermined value per square meter and within a bandwidth of 4 kHz. This limit protects terrestrial communication links (e.g., point-to-point radio systems).

At an elevation of 5° a power level of −155 dB is allowed which may linearly rise to −144 dB at an inclination of 25° and above (Figure 4.7.2).

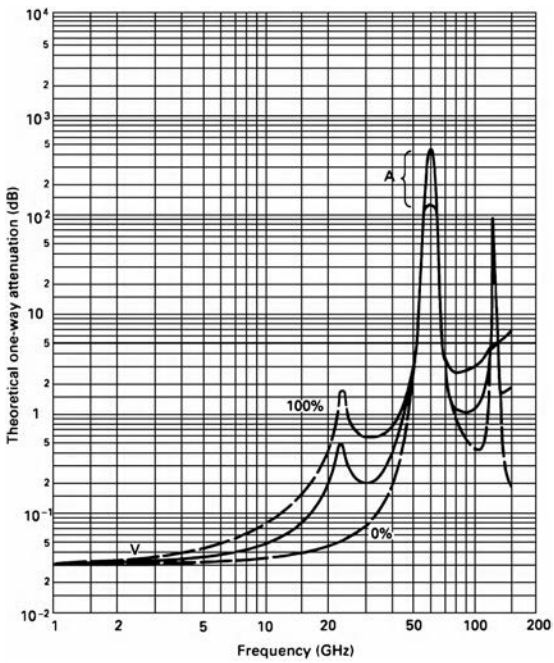


Figure 4.7.1: *Theoretical one-way attenuation for vertical orthogonal transit of a vertically polarized wave passing through the Earth's atmosphere (US standard atmosphere for 45° latitude). The solid line shows the attenuation gradient for medium water vapor content while the two dashed lines show the attenuations for 0% and 100% water vapor, respectively. The attenuation area marked with "A" reflects the error range at 60 GHz (Source: CCIR Report 390-4).*

4.7.3 Channel Capacity

Claude Shannon's channel capacity theorem [4.7.1] shows that a continuous band-limited signal has only a finite potential for error-free transmission of information.

In a band-limited transmission channel of bandwidth B [Hz] with an average thermal noise power of N [W] and Gaussian distribution, the **maximum error-free data rate** R [Hz] of an information unit with a signal power of S [W] can be expressed by the following relation:

$$R = B \cdot \log_2\left(1 + \frac{S}{N}\right) \quad [\text{bit/s}] \tag{4.7.1}$$

This is the limit of what is possible (comparable to the second law of thermodynamics) and serves to evaluate the efficiency of a real transmission system.

With the bandwidth predetermined, the only way to increase the possible data rate is to increase the signal power. As the channel capacity increases with the binary logarithm of the signal-to-noise ratio, the price for this approach can be very high.

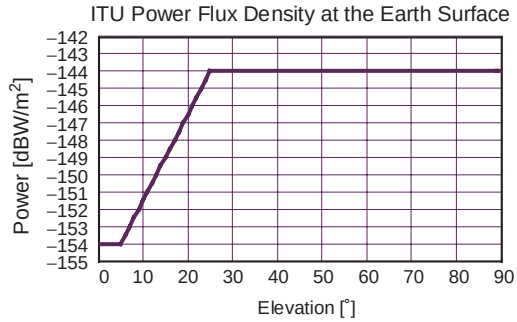


Figure 4.7.2: *Maximum power flux density allowed by ITU for an S-band signal which is incident on the Earth's surface.*

In most practical applications, the signal power is limited by fixed factors (e.g., efficiency of the transmitter, battery capacity). To approach the theoretically possible data rate, one tries to keep the **noise power** as small as possible.

By rearranging the Shannon formula one can find the **limit value of the channel capacity** for a constant signal power:

$$\frac{S}{N} = \frac{S}{k\,T\,B} \text{ and thus } B = \frac{S}{k\,T} \cdot \frac{N}{S} \tag{4.7.2}$$

where:
k = Boltzmann's constant, $k = 1.3806505 \cdot 10^{-23}$ J/K
T = Temperature [K].
Inserting B into the Shannon formula yields

$$R = \frac{S}{k\,T} \cdot \frac{N}{S} \cdot \log_2\left(1 + \frac{S}{N}\right) \quad \left[\text{bit/s}\right] \tag{4.7.3}$$

The limit of this function is

$$\lim_{\frac{S}{N} \to 0}\left(R\right) = 1.443\,\frac{S}{k\,T} \tag{4.7.4}$$

This result shows that with a given signal power S and a constant noise energy kT, the maximum error-free transmission rate can only be achieved with an **unlimited bandwidth** B ($S/N = 0$).

In real life this requirement cannot be fulfilled. But already with a signal-to-noise ratio of −10 dB one achieves 95% of the maximum possible data transmission rate.

Various modulation schemes such as code division multiple access (CDMA) distribute the available signal power over a large bandwidth and therefore come very close to the boundary identified by Shannon.

4.7.4 Antennas

The antenna is a passive (reciprocal) element used to couple an electromagnetic wave (existing in a waveguide or coaxial cable) to free space. Therefore, no difference exists between an antenna used for transmission and one used for reception in the following considerations.

4.7.4.1 Parabolic Reflector Antennae with High Directivity

A transmitting antenna for ease of imagination is considered. An electromagnetic induction field exists directly at the radiating element with rapidly decreasing field components. The subsequent electromagnetic radiation field can be divided into two regions:

- **Near-field** or **Fresnel region**
- **Far-field** or **Fraunhofer region.**

A parabolic antenna with a diameter D exhibits **diffraction phenomena** identical to those of a circular aperture exited by a plane wavefront. When measuring the diagram of an antenna one has to make sure that a **plane wavefront** from the transmitting antenna is arriving at the antenna to be measured. The changeover between these regions is at a distance r and depends on the diameter D of the antenna and on the operating frequency with wavelength λ:

$$r = 2\,D^2/\lambda \qquad [\text{m}] \tag{4.7.5}$$

Example
To accurately measure the diagram of a 3 m S-band parabolic antenna with a boresight transmitter, the transmitting antenna must be at a minimum distance of ≥ 135 m.

4.7.4.2 Gain of a Parabolic Antenna

When receiving an incident wave at the plane of the antenna with a **power flux density** [W/m^2], the received power is available at the antenna terminals. It is obvious that this power has a direct interrelationship with the physical area A of the antenna and its efficiency.

Every real antenna has a **main beam** in the direction of receiving (transmitting) and a number of (in most cases disturbing) **side lobes**.

The **gain** G (**directivity**) of an antenna in the directions θ (azimuth) and Φ (elevation) with an incident antenna diagram of $P(\theta, \Phi)$ is given by

$$G_A\left(\theta, \Phi\right) = \frac{P\left(\theta, \Phi\right)}{\frac{1}{4\,\pi}\int_0^{4\,\pi}\int_0^{4\,\pi} P\left(\theta, \Phi\right)\,\mathrm{d}\theta\,\mathrm{d}\Phi} \tag{4.7.6}$$

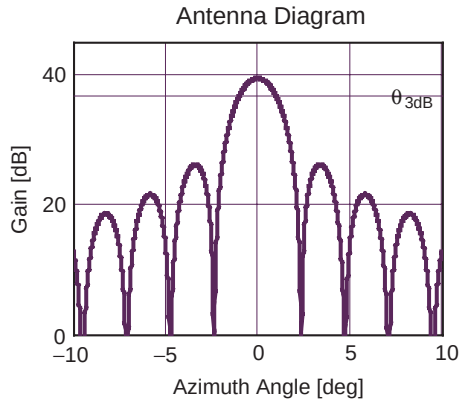


Figure 4.7.3: *Calculated antenna diagram for a 5 m parabolic antenna ($\eta = 0.6$) with a marker at the –3 dB beamwidth.*

In relation to the isotropic radiator[2] with $G = 1$ the gain of a real antenna is

$$G_A = \frac{4\,\pi\,A_{eff}}{\lambda^2} \quad \text{or} \quad G_A = 10\,\log_{10}(G) \quad [\text{dBi}] \qquad (4.7.7)$$

with $A_{eff} = A\eta$ the effective area of an antenna.

The value G denotes the directivity of the antenna in comparison to spherical wave propagation.

Real parabolic antennae have typical efficiencies η of 50–60% (monopulse autotracking receive/transmit or only receive/transmit). The main factors affecting the **efficiency** of an antenna are ohmic losses, reflection, and mechanical blockages caused by the mounting of the primary feed or by deviations of the parabolic dish from the ideal form.

The effective antenna area of an isotropic radiator is the area of a circle with one wavelength λ circumference.

4.7.4.3 Beamwidth of a Parabolic Antenna

A commonly used approximation for the beamwidth of a parabolic antenna is obtained by the following considerations and by assuming that the total energy is radiated within a uniform cone of angular radius θ [rad].

[2] An isotropic radiator is a hypothetical lossless point source which transmits the energy uniformly in all directions. The total radiated energy is present at the surface of the sphere $4\pi r^2$ with radius r.

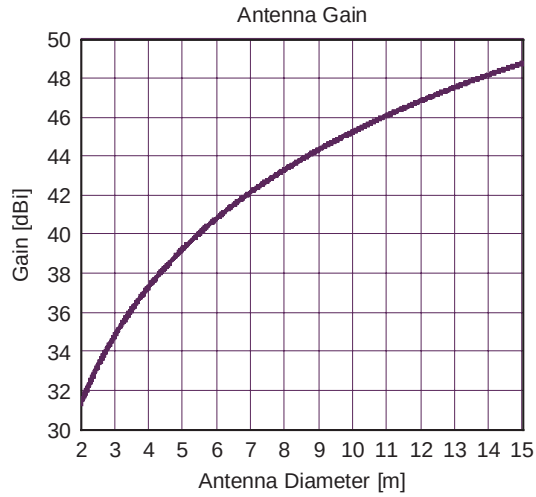


Figure 4.7.4: *Theoretical gain of an S-band parabolic antenna as a function of the antenna diameter for $\lambda = 0.133$ and $\eta = 0.6$.*

The antenna gain G_A describes the ratio of the total area of a sphere $4\pi r^2$ to the base area of the cone with its apex at the center of the sphere (the radius of the sphere is equal to the height of the cone):

$$G_A \approx \frac{4\,\pi\,r^2}{r^2\,\theta^2\,\pi} = \frac{4}{\theta^2} \qquad (\theta \ll 2\,\pi) \qquad (4.7.8)$$

Equation 4.7.8 provides an additional **definition of the gain of an antenna**: G_A denotes how often the solid angle of the antenna beam fits into 4π.

Using both equations for G_A and after some rearranging one gets the following approximation for the **half-power beamwidth** of an antenna (using $\eta = 100\%$ and D = antenna diameter):

$$2\,\theta = \frac{4\,\lambda}{\pi\,D} = 1.27\,\frac{\lambda}{D} \quad [\text{rad}] \qquad (4.7.9)$$

Example

For the theoretical gain of a 5 m S-band monopulse autotrack antenna ($\lambda = 0.133$ m) and beamwidth in degrees (Figure 4.7.3), the physical area of the antenna is $A = \pi D^2/4 = 19.64$ m^2. With an efficiency of $\eta = 0.6$ the effective area of the antenna is 11.78 m^2. Using these values the antenna gain $G_A = 10\,\log(8369) \approx 39.2$ dBi with a beamwidth of $2\theta \approx 1.9°$.

4.7.4.4 Polarization

The polarization of an electromagnetic wave is characterized by the **direction of the electric field component** (orthogonal to the direction of propagation). In satellite communications circular polarization is commonly used in addition to linear polarization (vertical or horizontal).

Depending on the direction of the field rotation (right hand or left hand looking into the direction of propagation) one differentiates between a **right hand circularly polarized** (RHCP) and a **left hand circularly polarized** (LHCP) wave. The axial ratio of an antenna is given by the following voltage relation:

$$AR = E_{max} / E_{min} \tag{4.7.10}$$

For an ideal circularly polarized wave $AR = 1$ while $AR > 1$ is obtained for an elliptic polarization. The sensitivity of a circularly polarized antenna to the orthogonal polarization is expressed by the **cross polarization discrimination** (XPD):

$$XPD = 20 \log_{10}\left(\frac{AR+1}{AR-1}\right) \quad [\text{dB}] \tag{4.7.11}$$

Modern communication antennae have a very high XPD value to be able to transmit or receive data simultaneously using the same frequency but with orthogonal polarization. Direct broadcasting satellites (DBSs) commonly use this feature to double their transmit channels in the respective frequency bands.

4.7.5 Thermal Noise

Gaussian thermal noise is omnipresent and plays an eminent role in modern telecommunications. Every transmitted signal is accompanied by a disturbing noise signal and each element in a transmission channel contributes with additional noise. Therefore, the signal-to-noise ratio of the entire communication link is deteriorated and it is necessary to consider all sources and reasons when specifying a communication systems. This assures that a useful signal arrives at the receiving end of the link.

Random movement of free electrons in a resistor R (which depends on its temperature) results in a **noise voltage** present at its terminals with a mean squared value of

$$\overline{v^2} = 4\,k\,T\,R\,\mathrm{d}f \tag{4.7.12}$$

where:
k = Boltzmann's constant,
$\mathrm{d}f$ = frequency [Hz],
T = temperature [K] of the resistor.
This consideration is valid up to frequencies $f \leq 4$ GHz. For higher frequencies the laws of quantum physics have to be applied. Taub and Schilling [4.7.2] extensively discuss the subject of noise.

The maximum available **mean noise power** N generated by a dissipative electrical two- pole network with a bandwidth of B is obtained by power matching the complex conjugate of the source impedance:

$$N = k\,T\,B \quad [\text{W}] \tag{4.7.13}$$

4.7.5.1 Antenna Temperature

To facilitate understanding of this notion we undertake the following visualized thought experiment. An antenna is located in a large closed room with wall temperatures T. When this system is in thermal equilibrium the room is filled with **black body radiation** of temperature T.

The output terminals of the antenna are matched with their characteristic impedance R to achieve power matching. With a true RMS voltmeter the voltage $\overline{v^2}$ is measured in the bandwidth B (f_1 to f_2). Rearranging Equation 4.7.12 and solving for T results in

$$T_{Ant} = \frac{\overline{v^2}}{4\,k\,R\,B} \tag{4.7.14}$$

and the mean available noise power at the antenna terminals is $N_A = kTB$ [W].

With real parabolic antennae with small beamwidths θ the **effective antenna temperature** T_{Ant} is equal to the temperature of the area at which the beam is pointing. At high elevation angles of the antenna, the temperature increase due to **cosmic noise** at frequencies $f < 1$ GHz has the highest contribution. At higher frequencies the contribution due to **atmospheric absorption** increases dramatically and therefore limits

the achievable system figure of merit. For frequencies below 10 GHz the actual antenna noise temperatures are a function of the elevation and are in the range of 150 to 15 K ($5° \le El \le 90°$).

A satellite receiving antenna pointing to the Earth (nadir) has at best a temperature of $T_{Ant} \approx 290$ K. The effective temperature can be much higher when flying over populated areas because of "human-made noise."

4.7.5.2 System Noise Temperature

The quality of a satellite ground station is determined by the antenna gain G_A relative to the isotropic radiator divided by the system noise temperature T_{sys}. The system noise temperature is composed of the antenna and receiver temperature and of additional temperature components caused by attenuation losses.

With the help of the following formula the **system noise temperature of the receiver chain** as shown in Figure 4.7.5 can be calculated:

$$T_{Sys} = T_A \cdot L_1 + (1 - L_1) \cdot 290 + (NF_1 - 1) \cdot 290 + \frac{(1 - L_2) \cdot 290}{g_1} + \frac{(NF_2 - 1) \cdot 290}{g_1 \cdot L_2} \quad (4.7.15)$$

For the mean physical cable temperature 290 K (≈ 17 °C) is assumed.

This formula shows that it is of utmost importance to keep the losses upstream of the preamplifier as low as possible in order to reduce their contribution to the overall temperature. Likewise, it is important to use a suitable preamplifier with a low-noise figure NF_1 and a sufficiently high amplification factor g_1 to compensate for subsequent cable losses.

Example (see Figure 4.7.5)

The temperature received by the antenna $T_A = 40$ K, the loss of the feed line $L_1 = 0.5$ dB. The preamplifier has a noise figure $NF_1 = 0.8$ dB with an amplification $g_1 = 38$ dB. The loss of the connecting cable to the telemetry receiver $L_2 = 0.5$ dB, while the receiver has a noise figure in S-band of $NF_2 = 12$ dB. Using these values results in a calculated system noise temperature of $T_{sys} = 132.7$ K. Due to the relatively high gain of the preamplifier, the temperature contribution of the cable $L_2 \le 1$ K. Assuming the gain $G_A = 40$ dB of the receiving antenna (with reference to the input of the preamplifier) the figure of merit of the system is

$$\frac{G_A}{T_{sys}} = 40 - 10 \log_{10}(132.7) \approx 18.8 \quad \left[\mathrm{dBK^{-1}}\right] \quad (4.7.16)$$

4.7.6 Modulation

4.7.6.1 Frequency Modulation

With frequency modulation (FM) the modulating signal changes the instantaneous carrier frequency by a value of Δf_T (**frequency deviation**). The absolute value of the change is only dependent on the amplitude of the modulation signal:

$$\Delta f_T = \beta \, f_S \quad (4.7.17)$$

The variable β is called the **modulation index**, and f_S denotes the highest frequency or the **bandwidth** of the modulation signal.

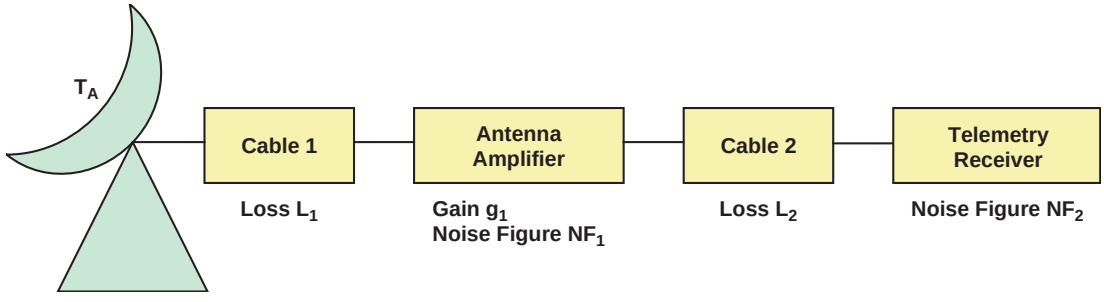


Figure 4.7.5: *Example of a ground station comprising a receiving antenna, a lossy cable ahead of the preamplifier and a second cable to connect the telemetry receiver.*

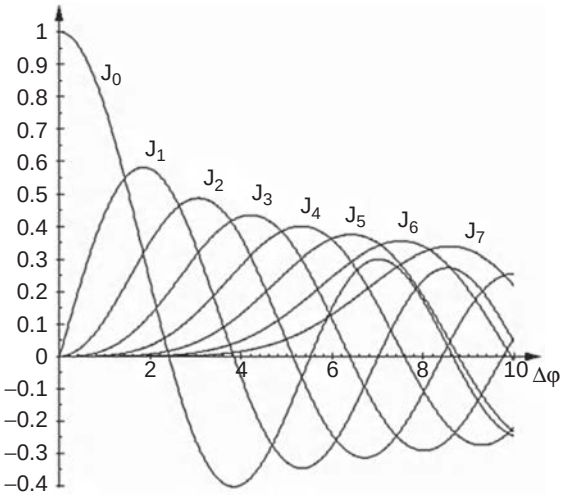


Figure 4.7.6: *First-order Bessel functions J_0 to J_7 as a function of β = 0 to 10.*

When modulating a carrier frequency with FM a number of sidebands are generated at a distance of $n \cdot \pm f_s$, whose amplitudes can be described by first-order **Bessel functions**. J_0 shows the amplitude of the carrier frequency while J_n show the amplitudes of the sidebands (Figure 4.7.6).

The bandwidth B of the carrier signal can be calculated with the well-known **Carson formula**

$$B = 2\left(\beta + 1\right) f_S \tag{4.7.18}$$

All sideband components outside of this bandwidth contain less than 1% of the transmitted energy and therefore can be neglected. It should be noted that the above considerations are only true for sinusoidal modulation signals. The bandwidth for FM with a band-limited signal with a Gaussian distribution can be obtained by solving the following formula for B:

$$\operatorname{erf}\left(\frac{B}{2\sqrt{2}\,\Delta f_{\mathrm{rms}}}\right) = 0.99 \tag{4.7.19}$$

After rearranging, B is given by

$$B = 2\sqrt{2} \cdot 1.825\,\Delta f_{\mathrm{rms}} = 5.16\,\Delta f_{\mathrm{rms}} \tag{4.7.20}$$

where Δf_{rms} is the Root Main Square (RMS) value of the frequency deviation.

4.7.6.2 Signal-to-Noise Ratio with Frequency Modulation

With FM the signal-to-noise ratio of the demodulated baseband signal S_o/N_o depends on β and on the ratio of received signal power S_i to noise power γf_S:

$$\frac{S_0}{N_0} = \frac{3}{2}\beta^2 \frac{S_i}{\gamma\, f_S} \tag{4.7.21}$$

where:
S_i = received signal power,
γ = spectral noise density (= kT),
f_S = modulation frequency.

Looking at this formula, one could derive that by increasing the phase deviation one could improve the resulting signal-to-noise ratio without limit. This would be true if one did not have to consider the well-known phenomenon known as the **threshold effect** for limited signal power. By definition the threshold effect is reached at a value of $S_i/\gamma f_S$ where the output signal-to-noise ratio S_0/N_0 deviates by 1 dB from its linear slope. To account for the threshold effect at low values of $S_i/\gamma f_S$ Equation 4.7.21 has to be augmented by the expression of the denominator in the following equation:

$$\frac{S_0}{N_0} = \frac{\frac{3}{2}\beta^2 \frac{S_i}{\gamma\, f_S}}{1 + \left(12\frac{\beta}{\pi}\right)\frac{S_i}{\gamma\, f_S}\, e^{-\frac{S_i}{\gamma\, B_{IF}}}} \tag{4.7.22}$$

where B_{IF} is the bandwidth of the intermittent-frequency filter.

The **spectral noise density** $G_{FM}(f)$ of a FM carrier signal is given (to a first approximation) by the following formula [4.7.3]:

$$G_{FM}(f) = \frac{\gamma f^2}{2 S_i} \prod\left(\frac{f}{B_T}\right) \tag{4.7.23}$$

with B_T the bandwidth of the carrier frequency. This expression shows that the spectral noise density increases with the square of the modulating frequency. This fact is especially disturbing when transmitting voice or music where the amplitudes of the higher

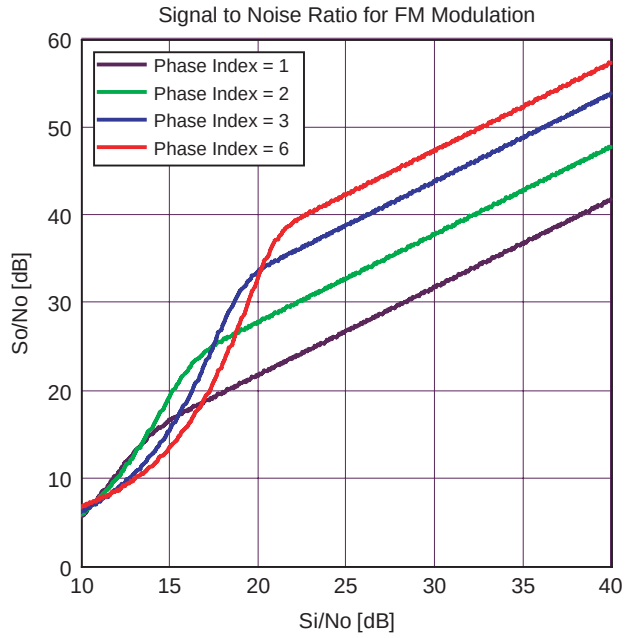


Figure 4.7.7: *Output signal-to-noise ratio as a function of the input signal-to-noise ratio and phase indices β=1,2,3,6. The onset of the threshold can be clearly seen and it is worth noting that for an input signal-to-noise ratio of 15 dB and β = 1 the output signal-to-noise ratio is higher than for β = 6.*

frequency contents of the signal are decreasing. The result is a degradation of the signal-to-noise ratio (cf. Figure 4.7.7).

The total **noise energy** N_D is obtained by integration of the noise density function over the bandwidth of the video filter $B_V = \pm f_V$:

$$N_D = \int_0^{B_V} G_{FM}(f)\,df = \frac{\gamma\, B_v^3}{3\, S_i} \qquad (4.7.24)$$

4.7.6.3 Pre-Emphasis and De-Emphasis

The received signal-to-noise power ratio in FM can be significantly improved by applying a linear accentuation (called a **pre-emphasis**, or an increase in the magnitude of the higher frequency signal components) to the modulating signal. Increasing the signal amplitudes by a high-pass filter with the transfer function $H_{Pr}(f)$ results in a larger frequency deviation for the higher frequency components of the signal.

After demodulation this pre-emphasis has to be equalized with the inverse transfer function (reduction of amplitude with a low-pass filter). This is necessary so that the original signal can be reproduced with high fidelity. Thus

$$H_{De}(f) = \frac{1}{H_{Pr}(f)} \qquad (4.7.25)$$

The transfer function of a simple *RC* low pass is given by

$$H_{De}(f) = \frac{1}{1 + j\,\dfrac{f}{f_1}} \quad \text{with} \quad f_1 = \frac{1}{2\,\pi\,RC} \qquad (4.7.26)$$

and the modified noise power spectral density is given by

$$G_{De}(f) = G_{FM}(f) \cdot \left|H_{De}(f)\right|^2 \qquad (4.7.27)$$

at the output of the **de-emphasis network**.

The improvement factor for FM is obtained after integration of the noise power spectral density function $G_{De}(f)$ over the bandwidth B_v:

$$\frac{N_{DPr}}{N_D} = \frac{3 f_1^3}{B_v^3}\left[\frac{B_v}{f_1} - \tan^{-1}\left(\frac{B_v}{f_1}\right)\right] \qquad (4.7.28)$$

Example

FM radio broadcasting stations in Europe use pre-emphasis networks with a time constant of $RC = 50\,\mu s$. This results in a corner frequency of $f_1 = 3.183\,kHz$ for the de-emphasis filter. All amplitudes of signals and noise above f_1 are attenuated by 6 dB/octave.

With monotransmission using an audio bandwidth of $B_v = 15\,kHz$ this approach results in an improvement of the signal-to-noise ratio of $\approx 10\,dB$.

For analog real-time transmission of high-resolution video information on microgravity experiments from on-board sounding rockets and satellites, high-efficiency FM S-band transmitters are used.

For this application the transfer function of the pre-emphasis network is defined by Consulting Committee on International Radio (CCIR) Recommendation 405-1 for PAL-G.

The transfer function of the de-emphasis network is given by the following formula:

$$H_{Pr}(f) = \sqrt{K}\sqrt{1+\left(\frac{f}{A}\right)^2}\cdot\left(\sqrt{1+\left(\frac{f}{B}\right)^2}\right)^{-1} \qquad (4.7.29)$$

In CCIR 405-1 for PAL-G, $K = 12.59$, $A = 1.565\,MHz$, $B = 0.313\,MHz$.

When transmitting video images with frequency contents up to 5 MHz, the average improvement in picture quality is 2 dB:

$$\frac{N_{DPr}}{N_D} = \frac{3\,K\,B^2}{A^2\,B_v^3} \times\left\{(A^2 - B^2)\left[B_v - B\tan^{-1}\left(\frac{B_v}{B}\right)\right] + \frac{B_v^3}{3}\right\} \qquad (4.7.30)$$

At first sight, an improvement of just 2 dB does not seem to be very much and one could ask if the effort in using premodulation and demodulation is worth the trouble. The necessary circuitry in the transmitter and the video receiver consists of only a few passive electronic components. To achieve the same picture quality without pre-emphasis and de-emphasis, one would need either to increase the transmitter power by a factor of 1.6 (e.g., 16 W instead of 10 W) or to use a larger antenna (e.g., 6.3 m instead of 5 m) for the reception of the transmitted signal.

After taking these figures into consideration it becomes rather clear that the use of premodulation and demodulation pays off in the end.

4.7.6.4 Phase Modulation (PM)

In contrast to FM, the modulating signal deflects the instantaneous phase of the carrier frequency by a value β (**modulation index**). The magnitude of this deviation in radians is only dependent on the amplitude of the modulating signal: $\beta = \Delta f_T / f_S$. In most practical applications a value of $\beta < 1.57$ rad is used to avoid ambiguity. With tone modulation with a sinusoidal signal, the resulting carrier signal is identical for FM and PM.

As a matter of interest, the **reception** of a PM carrier signal is possible with an FM receiver which has a linear *RC* low-pass (integrator) in series with its discriminator.

4.7.6.5 Frequency Shift Keying (FSK)

With **noncoherent FSK** the modulation frequency (mark and space) shifts the carrier frequency between two discrete frequency values. **Coherent FSK** (also binary FSK) has no phase change of the carrier frequency at the instant of frequency change.

4.7.6.6 Quadrature Phase Shift Keying (QPSK)

QPSK modulation is the result of simultaneously inverting the phase of two carrier frequencies which are orthogonal to each other. The possible phases of the carrier are:

45° ≡ binary 00, 135° ≡ binary 01,
225° ≡ binary 11, 315° ≡ binary 10.

In contrast to **binary phase shift keying** (BPSK) one can transmit 2 bits of information (one symbol) at the same time. With eight phase values one gets 8PSK modulation, and phase modulation (PM) is the (analog) PM with an infinite number of phase values.

4.7.6.7 Bit Error Rates for Different Modulation Schemes

In general the **complementary error function** erfc(x) is used to calculate the probability of the occurrence of a bit error in digital transmissions:

$$\mathrm{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-u^2}\, \mathrm{d}u \tag{4.7.31}$$

where $x = E_b/N_0$ (bit energy E_b to noise density N_0).

The boundary values for the complementary error function are 1 for $x = 0$ and 0 for $x = \infty$. Tabulated function values are found in [4.7.3].

For various selected modulations some formulas are listed below for calculating the bit error rate probability:

PSK or BPSK (or baseband)

$$P_e = \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{4.7.32}$$

Coherent Frequency Shift Keying (FSK)

$$P_e \approx \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{0,6\,\frac{E_b}{N_0}}\right) \tag{4.7.33}$$

FSK (noncoherent FSK)

$$P_e = \frac{1}{2}\,\mathrm{e}^{-\frac{1}{2}\frac{E_b}{N_0}} \tag{4.7.34}$$

BFSK (Binary Frequency Shift Keying)

$$P_e = \frac{1}{2}\,\mathrm{erfc}\left(\frac{1}{\sqrt{2}}\,\frac{E_b}{N_0}\right) \tag{4.7.35}$$

DPSK (Differential Phase Shift Keying)

$$P_e = \frac{1}{2}\,\mathrm{e}^{-\frac{E_b}{N_0}} \tag{4.7.36}$$

QPSK

$$P_e \approx \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\frac{1}{2}\,\frac{E_s}{N_0}}\right) \tag{4.7.37}$$

where E_S is the energy per symbol. With a given bit rate, transmission bandwidth and transmission power, QPSK has doubled the error rate of BPSK, but also doubled the transmission capacity.

PCM-RNRZ-L-FM (Randomized Nonreturn to Zero Level)

$$P_e = \frac{1}{2}\,\mathrm{erfc}\left[\frac{1}{2}\left(1 - \frac{\sin(2\pi\beta)}{2\pi\beta}\right)\frac{E_b}{N_0}\right] \tag{4.7.38}$$

15 bit pseudo noise (PN) code where β = modulation index.

Example

For BPSK modulation one gets a bit error rate probability of 10^{-6} for $E_b/N_0 = 10.53$ dB (Figure 4.7.8).

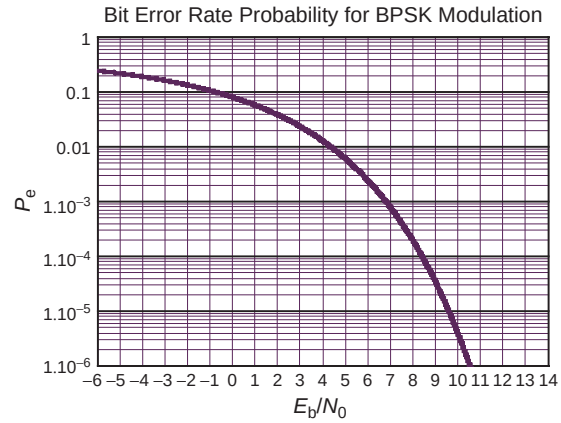


Figure 4.7.8: *Bit error rate P_e as a function of the ratio of the bit energy E_b to the noise energy N_0 for a transmission in baseband or with BPSK modulation.*

This is equivalent to one erroneous bit received in a million transmitted bits.

The formula indicates that on average every second bit is correct even with temporary total loss of transmission.

Depending on the nature of the satellite communication link, tolerable bit error rates are in the region of 10^{-4} to 10^{-7}.

When designing a system one has to take into account additional implementation losses (e.g., neighbor channel interference, receiver, bit synchronizer) of about 1 to 1.5 dB.

4.7.7 Pulse Code Modulation (PCM)

PCM denotes the **digital representation of an analog signal** for modulation. The analog value of a signal is sampled at regular time intervals and transformed to a digital value with a certain quantization (number of bits). With a finite quantization the analog value can only be represented with a residual **quantization error**.

In order to avoid **aliasing artifacts**,[3] it must be ensured that the sampling frequency is at least twice as high as the highest frequency contained in the analog signal. In many applications quantization of 8 to 16 bits is common. This is equivalent to a resolution of 40 to 0.15 mV assuming an analog signal with an amplitude of $10\,V_{pp}$:

$$f_S \geq 2\, f_N \qquad (4.7.39)$$

(the Nyquist–Shannon sampling theorem).

Stringing together all quantized values results in a serial stream of data consisting of a number of words. To enable the reverse process at the receiver end, a known **synchronization pattern** (sync words) is included in the serial bit stream. This entire data block is called a **PCM frame** (Figure 4.7.9).

[3] Frequency components of the wanted signal with $f > f_S/2$ are folded to a frequency band of $0 \leq f \leq f_S/2$. Therefore, the original signal has to be band limited before quantization to guarantee proper reconstruction of the original signal shape at the instant of digital-to-analog conversion at the receiver end.

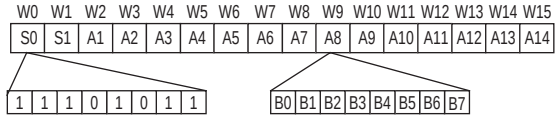


Figure 4.7.9: *PCM frame with two synchronization words (S0, S1) and 14 data words (analog channels A1 to A14) with a resolution of 8 bits/word. The synchronization words are predefined by the "Barker code" for 16 bits to EB90$_{hex}$.*

Example

For a bit rate of $f_b = 128$ kHz the resulting word rate is 16 kHz and the sampling frequency of each analog signal (A1 to A14) is 1 kHz. The maximum frequency content in the analog signals must be ≤ 500 Hz to avoid aliasing.

Coding Prior to Transmission

Depending on the application, a number of coding formats are used for the transmission of a PCM format. Some of the more common codes are shown in Figures 4.7.10 to 4.7.13.

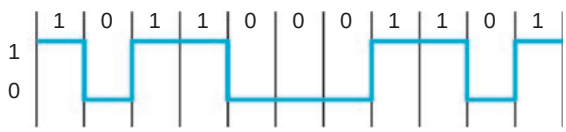


Figure 4.7.10: *NRZ-L.*

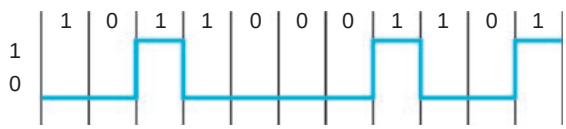


Figure 4.7.11: *NRZ-M.*

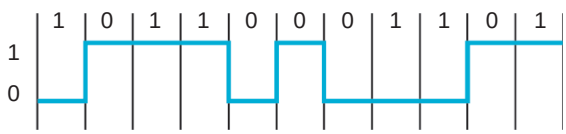


Figure 4.7.12: *NRZ-S.*

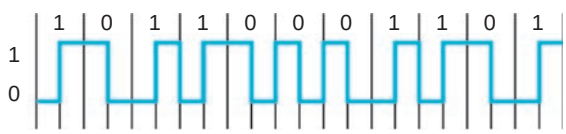


Figure 4.7.13: *Bi-Φ-L.*

The **nonreturn to zero level code** (NRZ-L code) transmits the digital bit value 0 with a zero level and the digital bit value 1 with a one level.

The **nonreturn to zero mark code** (NRZ-M code) transmits the digital bit value 0 with no change in level and the digital bit value 1 with a change in level.

The **nonreturn to zero space code** (NRZ-S code) transmits the digital bit value 0 with a change in level and the digital bit value 1 with no change in level.

The **Bi-Φ-L code** (also known as the Manchester code) is the logical "exclusive or" combination of the corresponding clock signal with the bit value.

This code produces a level change in the middle of each bit interval and therefore has no **DC component** and is self-clocking. This feature is obtained at the price of doubling the transmission bandwidth. For real applications the signal to be transmitted is band limited (premodulation filtered) with a six-pole Bessel filter with a cut-off frequency of $f_{-3\,\mathrm{dB}} = 1.5 f_\mathrm{b}$ prior to the modulation of the carrier signal.

A variety of additional codes exist (e.g., Miller or delay code modulation, Miller squared code) with most of them having no significant use in actual applications.

There is a high probability of having long trails of zeros and ones when using NRZ codes.

The resulting **DC level** of the modulating signal potentially can be a problem for the transmitter but also for the synchronization of the bit synchronizer at the receiving end. A possible remedy of this problem is to use a pseudo-random code (RNRZ-L) which can be obtained by implementing a linear feedback shift register.

For this case the maximum achievable pseudo-random length is $2^{15} - 1 = 32\,767$ bits. For automatic synchronization, the corresponding decoding logic needs 15 consecutive error-free bits (Figure 4.7.14). Single bit errors at the receiver entail two additional bit errors after the decoding process. In this case, the respective bit error rate for a given modulation has to be multiplied by a factor of 3.

To make clear the difference between NRZ-L and Bi-Φ-L, the normalized spectral energy densities for PR-PCM are depicted in Figure 4.7.15. It is obvious that for the transmission of an NRZ-L coded signal DC coupling is required, but frequency components above the actual bit frequency can be neglected. A Bi-Φ-L coded signal has no DC component but the frequency components are shifted toward the bit frequency, resulting in a doubling of the necessary transmission bandwidth.

The envelopes show a maximum of $f < 0.5 f_\mathrm{b}$ for NRZ-L and a maximum for $f \approx 0.75 f_\mathrm{b}$ for Bi-Φ-L.

Using the Carson formula for calculating the resulting bandwidth of the carrier frequency leads to an erroneous result when using PCM-RNRZ-L-FM due to the various frequencies at different amplitudes of the modulation signal (cf. Figure 4.7.16).

Empirical measurements with a spectrum analyzer show the following **bandwidths** as a function of β and of the frequency limits of the premodulation filter:

$$B = 1.16 f_\mathrm{b} \quad (\beta = 0.35, \quad \beta_\mathrm{Pr} = 0.7 f_\mathrm{b}) \tag{4.7.40}$$

$$B = 1.57 f_\mathrm{b} \quad (\beta = 0.40, \quad \beta_\mathrm{Pr} = 0.7 f_\mathrm{b}) \tag{4.7.41}$$

where B_Pr is the bandwidth of a four-pole low-pass filter with Bessel characteristics.

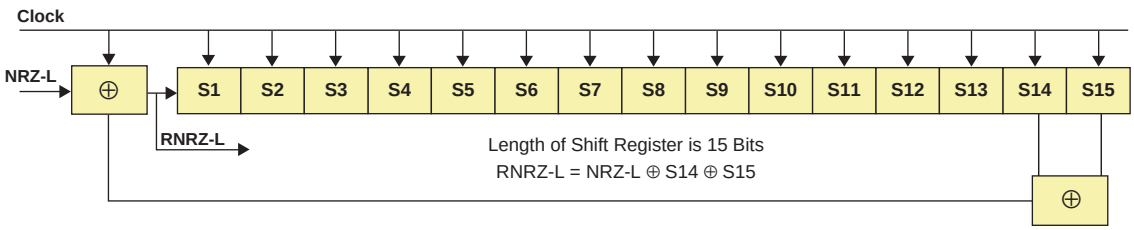


Figure 4.7.14: *Schematic block diagram of a 15 bit coding circuit as described by the IRIG 106 standard comprising a 15 bit shift register and two logic "exclusive or" gates.*

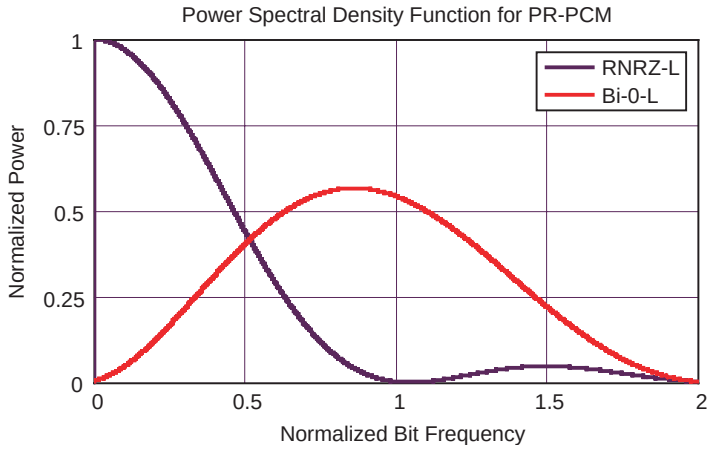


Figure 4.7.15: *Spectral energy density function for a NRZ-L and a Bi-Φ-L coded signal.*

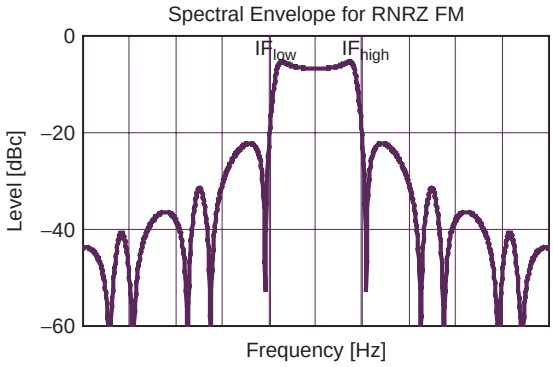


Figure 4.7.16: *Envelope of a RF carrier frequency modulated by a RNRZ-L signal (no premodulation filter used).*

Comprehensive information about various forms of modulation is given in the IRIG standard 106-07 [4.7.4].

4.7.8 Packet Telemetry

Modern telecommunications satellites transmit their data according to the CCSDS standard. The available transmission bandwidth is divided into a number of virtual channels. In this way the individual data sources (experiments or housekeeping) have the possibility to use the maximum possible **data rate of the telemetry** if needed. To safeguard the transmission quality the data is coded by a Reed–Solomon encoder block prior to a convolutional encoder. Using a 8 bit shift register with feedback, the data is serialized into pseudo-random NRZ-L format. The transmission of the data is within a transfer frame consisting of a fixed number of 8 bit words and preceded by synchronization bits.

The largest part of this **complex coding** and formatting is realized by special integrated hardware components which are qualified for space application. Dedicated programs are available for modern CPLDs (Complex Programmable Logic Devices) using the programming language Verilog HDL (Verilog Hardware Description Language) for implementation of the CCSDS transmission standard.

By using this coding scheme the theoretical transmission capacity of the channel is closely approached and a net gain of >5 dB is available for the link design.

4.7.9 Code Division Multiple Access (CDMA)

In contrast to FDMA (Frequency Division Multiple Access) trying to minimize the transmitting bandwidth, the transmitter power is spread over a large bandwidth (see Section 4.7.3) with CDMA where all users are transmitting their information at the same time in a **multiple access channel** with **spread spectrum modulation**. The information signal is multiplied (spread) with a **pseudo-random sequence** and the resulting transmission spectrum is similar to noise.

For demodulation of the signal, the receiver needs an exact copy of the pseudo-random sequence. A cross-correlator continuously compares it to the received signal and the output amplitude only reaches a maximum when an exact match occurs. All users are able to occupy the same data channel despite mutual interference because each user employs a unique and distinctive code.

Because of this tolerance to interference, the use of CDMA permits a reduction of the **channel distance** for satellites as well as for mobile telephones without significant loss of quality. Another advantage is the possibility to use smaller antennas without having to tolerate disturbances from neighboring satellites.

4.7.10 Coupling Networks

Quite often it is advantageous to operate two transmitters on one antenna system by selecting a coupling network with a transmission loss for each channel as low as possible and at the same time with a stopband attenuation as high as possible for the opposite channel. The most commonly used system is the **dividing/combining network** realized with high-quality bandpass sections and thus providing high selectivity for each channel. One major drawback of these units is the voltage overshoot within the individual filter sections and the associated risk of flashovers (corona effect at low pressures in the region of 10^3 to 10^2 Pa). To eliminate this risk it is necessary to design the unit to be absolutely hermetically sealed or to allow for complete outgassing (e.g., in satellites) of the system before its first use. Due to the relatively short exposure to vacuum, the employment of unsealed filter networks for use on sounding rockets is not possible.

For these applications the **phase line coupler** is an attractive solution. It consists of two 90° couplers (four poles) which are connected in series. One connection is direct while the other connection is implemented with a phase line.

When choosing the electrical length l_e of the cable so that the resulting phase for frequency f_1 is 0° and simultaneously 180° phase for frequency f_2, then both signals are directed (with low losses) to the antenna port "Ant" (see Figure 4.7.17):

$$l_e = (n + 0.5) \cdot \lambda_2 = n \cdot \lambda_1 \qquad (4.7.42)$$

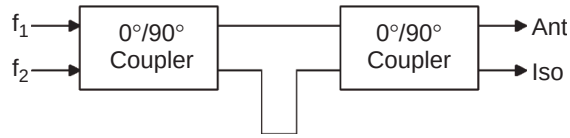


Figure 4.7.17: *Fundamental construction of a phase line coupler.*

with $n \in G$ and at the same time as small as possible. Deviations result in a phase error Φ and consequently in higher overall losses.

The signal power at the ports is

$$P_{\text{Ant}} = \frac{1}{4} \cdot \left|V_{f_1}\right|^2 \cdot a^4 \cdot \{1 + 2 \cdot d \cdot \cos[\Phi(f_1)] + d^2\} \qquad (4.7.43)$$

$$P_{\text{Iso}} = \frac{1}{4} \cdot \left|V_{f_1}\right|^2 \cdot a^4 \cdot \{1 - 2 \cdot d \cdot \cos[\Phi(f_1)] + d^2\} \qquad (4.7.44)$$

where:
a = transmission loss of one 90° coupler,
d = cable attenuation of the phase line.

Due to its small physical dimension, but also because of economic considerations, the phase line coupler was used for the BIRD satellite to couple one transmitter and one receiver with very low losses to a common antenna (see Figures 4.7.18 and 4.7.19).

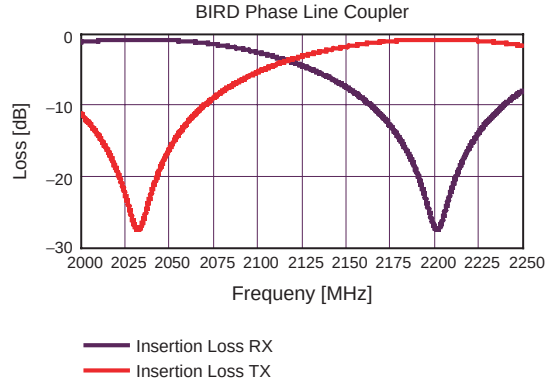


Figure 4.7.18: *Phase line coupler for BIRD. The resulting transmission loss is ≈ 0.8 dB with a stopband attenuation of ≈ 27 dB (a = 0.1 dB and d = 0.4 dB). The shallow curve of the passband attenuation shows the wide bandwidth of this coupling principle.*

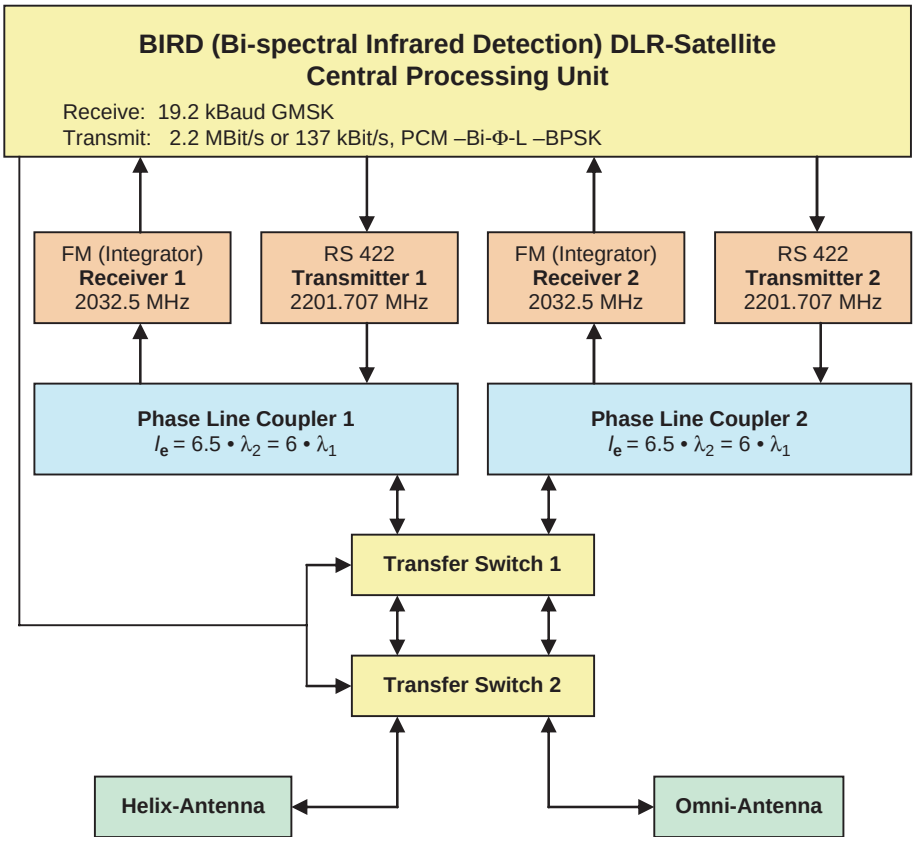


Figure 4.7.19: *Block diagram showing the redundant BIRD receive/transmit system comprising a high-gain antenna for transmission of spectral images and a low-gain antenna for low data rate, omnidirectional coverage. This assures access to the satellite during uncontrolled attitudes.*

4.7.11 Transmit and Receive System of the BIRD Satellite

The BIRD research satellite, developed and built by DLR, was also seen as a platform to **verify new space technology** apart from its scientific aims. This permitted the use of components for the communication system which had proven their use in sounding rocket projects. As normal sounding rocket payloads are only briefly exposed to space radiation (< 20 min), additional tests had to be performed to qualify these components since potential manufacturers in most cases could not provide the necessary information.

All components which could be affected by space radiation were tested with a total dose of > 15 krad at the test facilities of the Hahn-Meitner Institute in Berlin. This dose is equivalent to the calculated exposure of the BIRD satellite during its intended lifetime without taking into account the additional shielding effects of aluminum cases and structures. The nominal orbital altitude was established at 580 km and the DLR ground stations at Weilheim and Neustrelitz were chosen to operate the satellite.

4.7.11.1 Implementation

In order to compensate for single component failures, the transmitting and receiving system is designed to be fully redundant. In each case a transmitter and a receiver are coupled to a common antenna port with a passive low-loss (< 0.8 dB) phase line coupler. The actual physical cable length of the phase line is only ≈ 61 cm with $n = 6$ because of the large difference between transmit and receive frequencies. While both receivers are permanently in operation, the active transmitter can be selected. By means of the transfer relay (changeover switch) each transmit/receive unit can be connected either to the helix antenna ($G \geq 0$ dBi)

or to the omnidirectional antenna ($G \geq -10$ dBi). The second transfer switch is provided for redundancy reasons only.

The actual stopband attenuation of the phase line coupler is in the order of ≈ 27 dB and consequently not large enough for sufficient decoupling of the transmit and receive paths. With simple additional series resonance filters at the transmitter output $f_c = f_{RX}$ and at the receiver input $f_c = f_{TX}$ the decoupling can be improved to > 50 dB. This eliminates interfering broadband transmitter noise or a too large transmitter signal appearing at the receiver input.

The modulation of the S-band transmitters (f_{TX} = 2201.707 MHz) is PCM–Bi-Φ-L–BPSK for both possible bit rates (2.2 Mbit/s for image data and 137.5 kbit/s for housekeeping data).

The uplink modulation is Gaussian minimum shift keying (GMSK) PM with a constant bit rate of 19.2 kbit/s and forward error correction, implemented with the half-duplex packet data modem CMX900 from Consumer Microcircuits Limited.

True FM double-superheterodyne receivers with an additional linear integrator stage after the demodulator allow for correct reception of the PM data packets. In comparison to PM receivers it is much more economical to use true FM receivers for this application as this approach eliminates the need of sophisticated synchronization and verification logic for the lock on the received RF signal.

The frequency shift caused by the Doppler effect (≈ 50 kHz) of the received RF signal has to be accounted for in addition to the modulation when specifying the bandwidth of the receiver IF filter.

The achievable system sensitivity is quite satisfactory for commanding the BIRD satellite in LEO.

4.7.11.2 Telemetry Downlink Budget

The detailed link budget for the BIRD satellite is listed in Table 4.7.2. Worst case conditions have been used for all assumed values, for example allowing for reduced transmitter power at maximum temperature or for an adverse antenna diagram. For the receiving and transmitting stations, the actual values of an experimental 2 m ground station are used. No **additional margin** is required with these specifications as all operational ground stations have a much larger figure of merit for the uplink and downlink. The actual difference can be directly added to the contingency.

During the transmission of the scientific data, the on-board helix antenna is nadir oriented. When the satellite is rising above the horizon (0° elevation of the receiving station), the aspect angle onto the antenna diagram is ≈ 24° (elevation angle Φ of the helix antenna). For simplicity of calculation all values are used in their logarithmic form as multiplication reduces to addition and division to subtraction (see also Section 6.3.4.8).

To verify that the maximum permitted power flux density on the ground (ITU regulation) is not violated, Figures 4.7.2 and 4.7.20 are used. Assuming favorable conditions, the result shows that it is possible to receive image data from the BIRD satellite with a 2 m parabolic antenna above an elevation angle of ≥ 5°. For the uplink an EIRP of 45 dBW is sufficient (≈ 34 W RF power at the transmitting antenna).

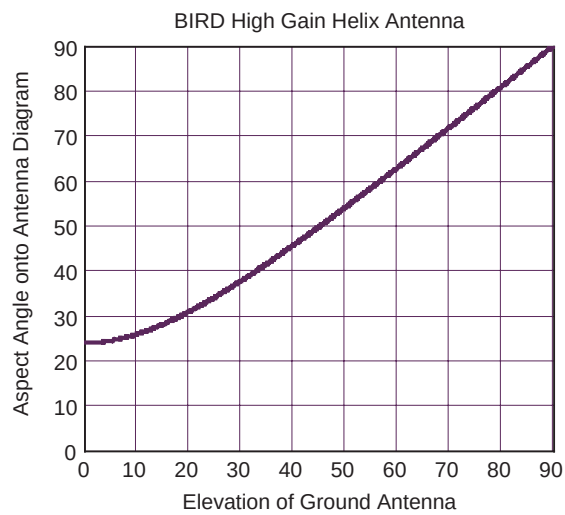


Figure 4.7.20: *BIRD high-gain helix antenna.*

Table 4.7.2: *Telemetry downlink budget for the BIRD satellite.*

On-board segment				
Transmitter power	P	5 W	7 dBW	
Transmitter frequency	f_{TX}	2201.7 MHz		Wavelength $\lambda = 0{,}136$ m
Modulation index	β	1.57 rad		Phase deviation $(\pi/2)\, f_b = 2.2$ MHz
On-board losses	L		1.5 dB	Coupling losses and cable losses
Antenna gain	G		0 dBi	Worst case (measured, nadir ± 66°)
EIRP (Equivalent Isotropic Radiated Power)	EIRP		5.5 dBW	EIRP = $P - L + G$
Space segment				
Minimum elevation	El	≥5°		Start of data reception
Orbital altitude	Alt	600 km		(Nominal 580 km)
Maximum distance	d	2328 km		For chosen elevation
Free space attenuation	FD		166.6 dB	$FD = 20 \log_{10} (4\pi d / \lambda)$
Atmospheric losses	L_{Atm}		0.5 dB	
Signal power at the antenna			−161.6 dBW	
Ground segment				
Figure of merit (actual value)	G/T		8 dB/K	Antenna gain of the 2 m ground station – system noise temperature
Boltzmann constant	k	$1{,}38 \cdot 10^{-23}$ W·s / K	228,6 dB W·s / K	
Signal-to-noise density ratio	S/kT		74.9 dBHz	$\frac{S}{kT} = \text{EIRP} - FD - L_{Atm} + \frac{G}{T} + 228{,}6$
Second IF filter bandwidth	B_{IF}	6 MHz	67.8 dBHz	BPSK-Modulation
Signal-to-noise ratio	S / N		7.2 dB	$S/N = S/kT - B_{IF}$
Result				
Bit energy to noise energy	E_b/N_0		11.5 dB	$S/N = S/kT - f_b$
Implementation losses	L_I		1 dB	Demodulator and bit synchronizer
Margin for $P_e \leq 10^{-6}$	Margin		0 dB	Required $E_b / N_0 \geq 10.53$ dB
Signal-to-noise ratio for carrier synchronization	S/N		7.2 dB	
Contingency			7.2 dB	Minimum of 0 required for “auto acquisition” of the BPSK demodulator

Bibliography

[4.7.1] Shannon, C.E. *Communication in the Presence of Noise.* Proceedings of the IRE, Vol. 37, 1949.

[4.7.2] Taub, H., Schilling, D.L. *Principles of Communication Systems.* New York: McGraw-Hill, 1971.

[4.7.3] Abramowitz, M., Stegun, I.A. *Handbook of Mathematical Functions*, Tenth Edition. National Bureau of Standards Applied Mathematical Series, 1972. (Free copy: http://www.math.hkbu.edu.hk/support/aands/toc.htm.)

[4.7.4] Range Commanders Council, Telemetry Group, Inter Range Instrumentation Group (IRIG). Standard 106-07, Part 1, September 2007. US Army White Sands Missile Range, New Mexico, 88002-5110.

[4.7.5] Carlson, A.B. *Communication Systems, An Introduction to Signals and Noise in Electrical Communication,* Second Edition. Kogakusha: McGraw-Hill, 1975.

[4.7.6] Viterbi, A.J. *Principles of Coherent Communication.* New York: McGraw-Hill, 1966.

[4.7.7] Consultative Committee for Space Data Systems (CCSDS). Recommendation CCSDS 101.0-B-6, Telemetry Channel Encoding. Blue Book Issue 6, October 2002 (http://www.ccsds.org).

[4.7.8] European Cooperation for Space Standardization (ECSS), Space Engineering, Communications Working Group, ESA Publications Division, ESTEC, Noordwijk, the Netherlands, *ECSS-E-ST-50C*, 2008.

[4.7.9] Gibbel, M. Thermal/Vacuum versus Thermal Atmospheric Testing of Spaceflight Electronic Assemblies. 16th Space Simulation Conference, Albuquerque, New Mexico, November 5–8, 1990, NASA Conference Publication 3096.

[4.7.10] Richaria, M. *Satellite Communication Systems.* New York: McGraw-Hill, 1999.

[4.7.11] Schwartz, M. *Information Transmission, Modulation and Noise*, Second Edition. New York: McGraw-Hill, 1970.

[4.7.12] Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, **27**, 1948.



5 Aspects of Human Space Flight

Helmut Luttmann

The following sections summarize aspects of human space flight. Putting humans into space can significantly improve the flexibility of research and operations as compared to robotic applications. Humans can interact with the systems and payloads, perform tasks which have not been defined in all their detail before a mission and communicate with scientists while carrying out the tasks of the mission. The presence of humans in space is a valuable resource which can be complemented by automation and robotic systems.

However, the consequence of this is, that significant technical support is required to allow humans to work and live there and to protect them against the hostile space environment.

The design drivers for human systems are robustness, reliability, safety, failure tolerance and ease of operation. Also, special functions such as caution and warning, safe haven and rescue systems must be considered. All systems must be designed so that they support the presence of astronauts. This includes supporting systems as well as payloads in terms of their access, interfaces, etc.

Astronauts must be accommodated in a pressurized shell with a supply of oxygen, removal of carbon dioxide, temperature and humidity control.

The design is significantly impacted by human factors: all systems must be tailored to the space and low-gravity environment. Available space on-board is limited. Systems must allow for efficient work with simple and safe operations. All materials must comply with environmental requirements (outgassing) and prevent any injury to the astronauts. The systems must comply with high safety and reliability standards.

For long-duration missions a resupply infrastructure is required to provide the astronauts with an atmosphere, food and clothing. Additional logistic tasks concern maintenance and repairs, including the delivery of spare parts, consumables, fuel, payloads, and waste disposal.

In the following sections of this chapter three main aspects of human space flight are discussed:

- General aspects of humans in space
- Environmental control and life support
- Rendezvous and docking.

The first section provides a general description of in-orbit activities and discusses the preparatory activities required on the ground to accomplish these tasks in space.

The second section describes the systems which support humans in space. Environmental control and life support includes all elements which enable the astronauts to work in a shirtsleeve environment, convenient for day-to-day presence. This function includes elements of atmospheric control and instrumentation, which prevent any danger to the astronauts, not least the caution and warning system. The section describes all the functions, the general layout, the system design and the components of the system.

The third section discusses in-orbit rendezvous and docking. All leading space agencies have decided to retain launcher systems that are unable to place an entire space station in orbit with a single flight. A

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9

project like the International Space Station requires multiple flights to deliver the planned elements and assemble them in orbit. In addition, regular resupply missions are required. This section discusses the principles of rendezvous and docking, the system design and the major components of the rendezvous and docking system as well as the underlying algorithms. The principles are illustrated using the automated transport vehicle (ATV) as an example.

5.1 Humans in Space

Hans Bolender, Reinhold Ewald and Klaus Wasserberg

In human space flight the astronauts (Western terminology), cosmonauts (Russian terminology) or taikonauts (Chinese terminology) are the focal point of the mission. This section describes the team's composition and allocation of duties (5.1.1), and summarizes its training (5.1.2) as well as the requisite infrastructure (5.1.3). As is the case in aviation, this team is frequently referred to as a crew, and the same thing will be done in this book since the term is now in common usage.

5.1.1 The Crew of the International Space Station

The International Space Station (ISS) will, for the foreseeable future, be the only **permanently occupied station** in space. ISS will therefore be used as the starting point for describing the composition of the crew and the tasks of its members. Daily life on-board a space vehicle and the methods of working together as a team will also be illustrated using the ISS as an example.

5.1.1.1 Composition and Selection of the Team

The crews of the ISS consisted of up to three members during the station assembly period. After the loss of the Space Shuttle *Columbia*, the crew was even reduced to two astronauts. This was done to conserve resources, which could only be brought to the station by the Progress and Soyuz transporters. This remained the situation during the period from February 2003 to July 2006. However, the size of the space station and the multitude of servicing tasks require a minimum crew of two astronauts just for ongoing maintenance. For sustained work on the scientific experiments, the experience of the last few years has shown that a crew of at least three astronauts is required.

The **composition of each crew** is negotiated by an international body, the Multilateral Crew Operation Panel (MCOP), made up of the partners participating in the space station, namely the USA, Russia, Europe (represented by ESA), the Canadian Space Agency (CSA) and Japan's Space Agency (JAXA). This panel authorizes ISS training for the astronauts and nominates primary and backup crews. ISS crews are assigned to **expeditions**. These are core teams of at least two people who fly to and return from the station together in the same space vehicle. Because of the reduced Space Shuttle flight plan, the expeditions continue to fly to the space station and back in the Soyuz capsule. The composition of such crews is governed by the principle that the major partners, the USA and Russia, are always represented by one crew member, the reason being that they are responsible for their respective station segments.

With the arrival of a significant contribution in the form of a pressurized module (USA, Russia, Europe, Japan) or a station tool like the Space Station Robotic Manipulating System (SSRMS, a resource provided by Canada), the spirit of **ISS partnership** became a reality, and the station's resources are at the disposal of new partners in proportion to their contribution. The same applies to crew time and the right to have a partner's own astronauts join an expedition crew. Putting this right into practice is the responsibility of the MCOP.

With the arrival of the Columbus laboratory module Europe earned approximately 8.5% of the **resource rights** to use the US segment and can plan on having its own (ESA) astronauts join expeditions. Depending on the planning, every two to three years ESA astronauts will be sent to the ISS for the duration of an expedition.

SYSTEMS/OPERATIONS	System Qualifications & Responsibilities				СИСТЕМЫ/ОПЕРАЦИИ
	Segment	Astronaut A	Astronaut B	Astronaut C	
Complex Operations					**Комплексные операции**
Node 2 Vestibule Outfitting	ISS	S	S	-	
Columbus Vestibule Outfitting	ISS	S	-	S(P)	
ELM-PS Vestibule Outfitting	ISS	S	-	-	
JEM Vestibule Outfitting	ISS	S	-	-	
Command & Data Group (CDG)					**Группа управления бортовой аппаратурой (CDG)**
Operations Local Area Network (Ops LAN)	ISS	S (P)	O	O	Локальная вычислительная сеть по операциям (Ops LAN)
Command & Data Handling (CDH)	USOS	S (P)	O	O	Система выдачи команд и обработки данных (C&DH)
Onboard Computer System (БВС)	ROS	O	S	O	Бортовая вычислительная система (БВС)
Onboard Equipment Control System (СУБК/А)	ROS	O	S	O	Система управления бортовым комплексом/бортовой аппаратурой (СУБК/А)
Data Management System (DMS)	EOS	O	U	S	
Command & Data Handling (CDH)	JOS	O	U	-	

Figure 5.1.1: *Excerpt from a crew qualification and responsibilities matrix (CQRM): S, Specialist; O, Operator; U, User; P, Prime; USOS, US On-board System; ROS' Russian On-board System; EOS, European On-board System; JOS, Japanese On-board System.*

5.1.1.2 Tasks of the ISS Crew

The ISS considerably surpasses the former MIR space station in mass, living space and complexity. For the crew on-board, this complexity means numerous special tasks which are identified in the so-called **crew qualification and responsibilities matrix** (CQRM) (Figure 5.1.1). In order to equip the crew with all the required qualifications in the available training time, compromises must be made as to the depth of knowledge communicated to each individual crew member. This is an evolution of the customary concept for distributing tasks on-board the Space Shuttle. In this vehicle the **flight tasks**[1] are assigned to the commander (CDR) and the pilot (PLT), with the support of **mission specialists** who are located on the flight and mid decks. The minimum number needed to perform all the tasks associated with the technical systems of the Space Shuttle is five. If the flight involves a complex payload, two to a maximum of three **specialists** are added. Their training is limited to the safe transport of and work associated with f.e. scientific experiments and to technical testing. Of late, mission specialists have also been taking over these payload tasks and, accordingly, flights with payload specialists have now stopped. Another group of Space Shuttle crew members are **astronauts with special responsibilities**, such as astronaut educators. These are astronauts whose main responsibility is in the area of out reach.

In the **Russian system** the Soyuz commander and the Soyuz flight engineer (FE) handle all flight tasks. The third seat in the narrow Soyuz capsule is occupied either by a "space flight participant" or tourist, or by a

[1] The Space Shuttle is both an orbiting space vehicle and an aircraft (during its propulsion-free return to Earth). Therefore, the division of tasks in the cockpit is very similar to that in an aircraft. This is also true of the seating positions for the commander (left) and the pilot (right). The position of the access hatch on the port (left) side is also analogous to aircraft.

professional astronaut from a space agency, normally an additional ISS crew member. Once on-board the ISS, the Soyuz commander and the Soyuz flight engineer either become part of the new expedition crew or return to Earth with the short-term visitor. This was the case for Soyuz replacement missions – the so-called "taxi flights."

For the space station crew two positions are defined: the **station commander** (ISS CDR) and the **flight engineer** (FE-1, FE-2). The authority of the ISS CDR is described in the Space Station Code of Conduct[2] and is legally binding for all designated ISS crew members for the duration of their training and flight. In turn the ISS CDR is responsible to the ISS flight director at Mission Control Centre in Houston for all aspects associated with the safe and successful realization of the mission. Although all partner control centers are basically responsible for their own space station contribution, the special role of the ISS flight director is specifically emphasized with respect to the ISS crew. This is a reflection of the major contribution of the USA to the ISS. Both positions, ISS commander and flight engineers, share the routine and special tasks that are to be carried out during the expedition. (See the example for a CQRM shown in Figure 5.1.1.)

All the tasks to be dealt with during an expedition are defined and negotiated long before the event and specified in various documents.[3] The result is a collection of major objectives for each expedition as well as different training requirements. If a crew member undergoes extensive training as a specialist for particularly difficult repairs or overhauls, it must be guaranteed at the same time that the necessary replacement parts for this repair will be sent to the space station in time for the work to be carried out.

The underlying concept of the ISS is that of a **laboratory** providing all partners with free, unlimited access to the laboratory facilities, tools and crew in proportion to their contribution. In principle no distinction is made for a planned experiment as to whether it is carried out by a Japanese or a European astronaut in the Columbus module. In practice, however, it turns out that, despite the mechanical coupling of the station segments, the station is clearly separated into a Russian part and a part which is operated by the USA and its partners Canada, Europe and Japan. The maintenance of the Russian segment alone requires the entire crew time of at least one cosmonaut. The US astronauts are primarily trained for work, maintenance and the repair of the systems in the US orbital segment and accordingly are assigned these tasks. Despite such restrictions the ISS is the biggest and most successful international civilian project ever realized.

5.1.1.3 Daily Routine of the ISS Crew

The daily routine on-board the ISS is the result of a long planning process. Parallel to crew training, a catalog of tasks to be accomplished week by week during the expedition is defined by the international planning teams. When the crew is launched, this collection of tasks (called the **On-Orbit Summary**, OOS) is available along with a more detailed breakdown of the individual tasks for each day of the following week. The daily plan, which the crew members on-board also see on their computers, is in final form up to four or five days in advance, but can be modified in the course of the various on-board activities. The final plan is sent to the station the evening before, after it has been agreed by the planning groups. Changes that occur afterwards are discussed with the crew in daily planning conferences. These planning conferences take place each morning and evening and are fixed points in the daily routine. Between waking up at 06:00[4] and the planning conference two hours later,

2 The ISS Crew Code of Conduct was negotiated in the MCOP and approved by all partner agencies. Besides the question of authority on-board, it regulates issues of liability, applicable law, legal jurisdiction back on Earth, ethical aspects and the obligation of expedition members not to use their position as an ISS crew member for personal gain.

3 The most important documents are the Tactical Plan (TP) and the Increment Description and Requirements Document (IDRD), which NASA negotiates with the partners for each increment (the duration of an expedition on-board the space station). The priorities for complex tasks are similarly prescribed.

4 The entire crew follows a common schedule without work shifts. The clocks on-board the ISS are set to Greenwich Mean Time (GMT), which also does not change during the summer months. This is a compromise to make a joint planning routine possible for the participating control centers. Because of the 16 Earth rotations per day, day and night on Earth naturally play no role on-board the ISS in giving a subjective sense of time.

the astronauts have free time during which they can acquaint themselves with the daily plan.

After the 15 minute planning conference, in which all crew members and control centers participate, the astronauts follow their individual **time line** until a common meal at noon. The same is true of the afternoon up to the planning conference in the evening. During the morning and afternoon, two and a half hours of fitness training is also scheduled. This includes the time necessary to change clothes and for personal hygiene. After the day's work the crew has time for an evening meal and to prepare for the following day. Because of the many replacement parts, experiments and consumables stored everywhere onboard, it is advisable to collect in advance whatever is going to be needed for the following day. This is also the time to answer emails, to make any notes that there was no time for during the day, or to telephone friends and family. The crew can make **telephone calls** via a broadband communication channel to Earth via relay satellites. This is a huge improvement compared to the era of the MIR space station, when communication was only possible during flights over Soviet ground stations. The working day is limited to a maximum of eight hours, including preparation, follow-up and sport activities. An international team of **ISS crew surgeons** monitors adherence to this time schedule. Interruptions to sleep, due to such things as unavoidable work on experiments, must be compensated with the equivalent free time. Nevertheless, members of long-duration expeditions reported that they actually never felt off duty, which is understandable in view of the unusual workplace and challenging environment.

Experience with the MIR space station has shown that **overworking** the crew is a constant risk. Therefore, in contrast to short Space Shuttle flights with spacelab, there are no work shifts for the European Columbus module. While the crew sleeps, the ground stations take control of the station via telemetry and telecommanding. Acoustic signals and warnings are suppressed and problem solving is postponed to the next morning wherever possible. On Saturdays, in general, two hours at most is spent on technical and research work. Cleaning activities are also on the program. Sunday is a free day and the time is used for video conferences with family and friends.

5.1.1.4 The ISS Crew as a Working Unit

Two criteria were mentioned above for arriving at **composition of an ISS expedition crew**. The first determines which astronauts from the various partner organizations participate in accordance to their organization's share in the ISS. The second criterion is that the crew as a whole must cover all the necessary qualifications as specified in the CQRM. Simply bringing experts together to form a team that has to subsequently dedicate itself in almost complete isolation to a set of tasks for a period of time does not guarantee success. Much empirical information is available on comparable situations, such as teams which spend the winter at polar stations, teams of mountain climbers or, in general, expedition teams in remote areas of the world. Discussions about the right crew composition include all the usual criteria such as the optimal number,[5] whether the crew should be mixed or of the same gender,[6] and hierarchy and seniority issues.[7] There is no reliable formula which guarantees a harmonious and successful crew composition. Nevertheless, there are a few characteristics which have supported especially successful ISS crews in their work.

The Astronaut Selection Process

Current astronauts in the US, European and Japanese teams were not recruited specifically for long-term missions on board the ISS, but for programs which were relevant at the time of their recruitment. Examples are the Space Shuttle program with its brief flights (also with Spacelab), or the Columbus program with its emphasis on laboratory work, but also the later

5 Because of the choice of the Soyuz capsule as a rescue vehicle and the capacity of the ISS, expedition crews of only two, three or later six (which presupposes two docked Soyuz capsules) are conceivable.

6 Several female astronauts who completed long-term expeditions have in the meantime demonstrated that in view of the work performed on-board, all the evidence supports having a mixed crew. However, intrusive attention from the media is to be expected.

7 It should be noted that the ISS serves to fulfill civilian scientific and technical objectives and normally manages quite well with a fairly shallow hierarchy. Only in emergencies and when responsibility has to be assumed for safety on-board is a commanding presence definitely required.

abandoned Hermes space glider project. Only Russia has not changed its recruitment criteria and the composition of its crews based on its experience with the MIR station. The relatively new requirements for astronauts are symptomatic of the difficulties NASA had in finding candidates for the expeditions on-board the MIR station. Even today, the long training period for a flight to the ISS and the special tasks on-board the ISS are far more demanding than what is required for a short mission. In addition, selection is typically influenced more by national programs and priorities than by ability to work in an international environment.

The Cultural Component

The Russian–US collaboration on the MIR station showed how important it is for the team to have a good command of at least one **common language**. The dominant language on-board MIR was Russian. But even today, when English is the main language used on the ISS, the team members are more in tune with each other if the Americans and Russians understand each other and each other's station-to-ground communications. The solution to this is an intensive and lengthy linguistic immersion in the other country before the flight. This can, however, conflict with the training regime and the wish to be with one's family.

However, overcoming the language barrier is not enough. Handling the technology, the repair strategy, on-board equipment storage and the scarce amounts of free time, which the control centers allow the crew members, are examples of different approaches taken by the west and the east with respect to operating a space station. Success depends on being extremely flexible with regard to **work methodologies**, and recognizing and anticipating the different leadership styles of the various control centers. Training units on crew resource management are a useful precautionary measure. They help to make evident the borders between a regulating hierarchy and personal judgment, which is similar to the situation in an aircraft cockpit.

The Inner Factor

During long-term missions the initial **euphoric phase**, which lasts about three to four weeks, is replaced by a **routine phase**. This is characterized by repetitive work, situations where an astronaut makes a mistake for the first time, and especially by the monotony of daily routine, which is governed by the constant necessity to stay physically fit and to adhere to a strict schedule. Sensory and communicative limitations also play a role here. This phase can be interrupted by external factors such as the arrival of a Progress/ATV transport vehicle or by a planned space walk. The ground control centers also watch for signs of moodiness amongst crew members and take corrective action since these symptoms increase the likelihood and frequency of operational errors. Such moodiness disappears as the flight comes to an end and activities associated with packing up and documentation begin, which require a high level of concentration. At this stage positive anticipation of the return to Earth dominates. Decisive for success, however, is the will of each individual member of a long-term crew to regard every moment of the flight as a contributing factor to the overall success of the entire mission, and at the same time to regard this as an integral part of their freely chosen career. The expectation of some kind of reward after a flight has not proven to be a worthwhile incentive for dealing with the sometimes unexpectedly frustrating monotony of the tasks on-board compared to idealized expectations. So the search for astronauts having the "right stuff" during the early days of space flight has now become a search for the "right staff" – for a crew which is technically excellent but which also brings along sufficient inner motivation to ensure a cohesive team capable of coping with the difficult phases of the flight.

5.1.2 Astronaut Training

Before flying to space astronauts have to undergo an intensive training program lasting several years. Flight preparation and training take up more time than the actual flight itself.

In the course of international space flight cooperation, astronaut training takes place at the various training centers of the space agencies participating in the programs. For the ISS program, astronaut training has evolved into a complex, global project with training centers located at Houston (USA), Star City near Moscow (Russia), Tsukuba near Tokyo (Japan) and at ESA's European Astronaut Centre (EAC) in Cologne, Germany.

Training programs are multilaterally coordinated and widely integrated to guarantee a highly effective and efficient training program for the astronauts and cosmonauts involved.

5.1.2.1 Astronaut Training Program

The astronaut training program is composed of three phases. Training phases may vary depending on the flight duration (short-term versus long-duration missions), the transport system (e.g., the Space Shuttle or Soyuz) and depending on the flight program and task distribution (e.g., scientific objectives, pilot, board engineer).

This section provides a detailed overview of the three-stage astronaut training concept presently in place for the ISS program (Figure 5.1.2). It is agreed upon and jointly implemented by the training organizations of all ISS partners.

Basic Training

Each space agency implements basic training for its own, newly recruited astronaut candidates. Basic training has a duration of 12 to 18 months and comprises more than 1200 training hours. The main objectives are:

1. To provide new astronaut candidates with a good overview of their future field of work.
2. To deliver basic knowledge and skills in various fields of human space flight which are particularly significant for a subsequent career as an astronaut.
3. To balance different levels of knowledge and education in the astronaut corps deriving from the different professional backgrounds of the individual candidates (e.g., scientists, test pilots, doctors, etc.).

Basic training consists of four instructional blocks (A–D), as follows.

A: The **general introduction** provides preliminary orientation. Astronaut candidates receive information on the different space agencies and space industry as well as on major historical human and unmanned and space programs, space law and all important current space programs and international cooperation. This training block also includes visits to the various locations of each candidate's own space agency as well as short stays at the training centers of the international partner organizations.

B: The training block of **fundamentals of human space flight** comprises selected lessons on space flight and aerospace engineering, aerodynamics, propulsion,

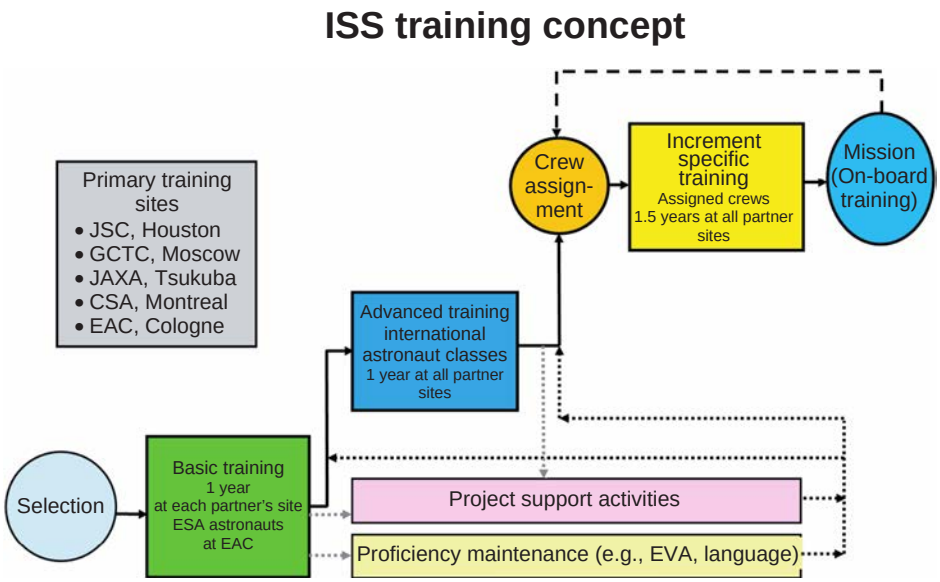


Figure 5.1.2: *The three stages of astronaut training in the ISS program.*

orbital mechanics, guidance, navigation and control, structures and materials, thermal control, electrical engineering, computer engineering, life science and biomedical science, material science, fluid dynamics, Earth and space sciences.

C: This course block on the **systems and operations of the ISS** provides an overview of important ISS documentation including the international agreements and resulting multilateral task sharing. Furthermore, the major on-board and ground systems of the ISS as well as the basic operational concepts are introduced. It also imparts insight into the systems and operations of the participating space transport vehicles (the Space Shuttle, Soyuz, Progress, ATV and HTV). Each of the following systems is introduced:

Ground systems and infrastructure:
- Launch facilities and launch preparation
- Development and test centers
- Control centers including mission operations concepts.

On-board systems:
- Principle mission concepts and spacecraft systems
- Structures and materials (including the ISS design philosophy and concept)
- Motion control
- Thermal control
- Electrical power
- Command and data handling
- Communication and tracking
- Environmental control and life support
- Extravehicular activity (EVA)
- Habitation and crew equipment
- Integrated medical systems
- Photo/TV systems
- Maintenance and repair
- Payloads.

D: The last training block is dedicated to **training of special skills**, which are of relevance for the future career of the astronauts. It involves the training of special psycho-motor, cognitive and psychological skills. The main topics are:
- Generic robotic training (GRT)
- SCUBA diving and EVA familiarization
- Flight training and parachute jumps
- Human behavior and performance (HBP) training for long-duration flights
- Russian language training.

The basic training minimum requirements are laid down in a multilaterally agreed document ensuring that all international astronauts have a comparable set of knowledge and skills before entering the second phase of the astronaut training program.

Advanced Training

Advanced training is conducted in **international astronaut classes**. The ISS partners nominate astronauts for these classes. These astronauts undergo the advanced training program and can afterward be nominated for an ISS space flight mission. Advanced training lasts about a year and takes place at the five training centers of the ISS partners. Each ISS partner is in charge of providing the astronaut training on all those components and systems which its own agency contributes to the ISS program. For example, the ESA is responsible for:
- Columbus research module systems and operations training
- Automated transfer vehicle (ATV) training
- Payload training for the ESA payload facilities (i.e., Biolab Module, European Drawer Rack (EDR), European Physiology Module (EPM), Fluid Science Lab (FSL) and several smaller ESA payload facilities and experiments).

The main objective of advanced training is to provide detailed knowledge and skills in operating and maintaining all ISS on-board and transport systems and subsystems. The training concentrates on generic and recurrent standard tasks and operations which are needed during any mission to the ISS. Mission-specific content such as ISS assembly tasks, special repair activities, etc., are not part of advanced training. As for basic training, advanced training concludes with the attainment of a set of qualification levels, which are defined as a prerequisite for the next training phase (see also Section 5.1.2.3). After having successfully concluded advanced training, astronauts are eligible for a space flight mission.

Increment Specific Training

Increment specific or mission training marks the last phase of the preparatory program prior to a space flight. It starts after the assignment of the prime and backup crew for an ISS mission. The 18-month increment specific training template is tailored to the specific flight program requirements. Based on the generic training content received during advanced training, the crew will be trained in this last training phase to conduct all flight-specific tasks which have been manifested for the flight. Generally the training program includes:

- ISS assembly tasks like the docking, checking out and commissioning of new modules
- Preplanned maintenance, upgrade and repair tasks (like filter exchanges, S/W upgrades or H/W replacement or repairs)
- Special preplanned EVAs or robotics activities
- Flight-specific scientific and technology experiments.

Training of complex operational tasks such as planned EVAs requires many hours. It comprises several repetitions of the whole EVA procedure to help the astronauts memorize the complex sequence of activities, to optimize the coordination and communication between the two space walkers, thus providing for maximum efficiency and safety.

Toward the end of this training phase the training focuses on potential off-nominal situations, including failure analysis and recovery activities. In addition, the astronauts participate in simulations of complex on-orbit activities, thus applying everything they learned within a realistic operational scenario. Another aspect is the simulation of ISS emergency situations such as on-board fire or rapid decompression. All the on-board crew and the ground control teams are involved in these simulations to train for the required efficient and well-coordinated reactions in a complex and time-critical environment.

At the end of increment specific training, the training organizations sign a "Certification of Flight Readiness" (CoFR) confirming that the primary and the backup crew successfully concluded the training and that – from a training perspective – they are prepared for and capable of conducting the flight program as planned.

The three phases of the astronaut training program as shown in Figure 5.1.2 present an ideal flow which is not always followed continuously. At the end of every training phase, astronauts may be temporarily assigned to so-called collateral duties, such as the support of new development programs, or assume certain team lead functions, before entering the next training phase.

During collateral duty times, astronauts receive limited refresher and proficiency training in order to retain skills and to maintain their qualifications and certificates like pilot or diving licenses.

On-board Training

Also on-board the ISS, it is necessary to continue training to a limited extent. Crews are regularly involved in emergency drills and undergo refresher training in order to maintain certain complex psycho-motor skills, such as those needed to operate the ISS robotic arm.

About two weeks prior to the rendezvous and docking of the ATV cargo vessel with the station, the astronauts on-board the station use an ATV on-board simulator to exercise once again the monitoring tasks and appropriate reactions to off-nominal situations which may be required from the crew during the forthcoming rendezvous and docking maneuver. The results of these simulations are downlinked to the ATV instructors on the ground, who evaluate the training status of the crew. If needed, additional failure cases can be added to a simulation to focus on the training of specific tasks or skills.

5.1.2.2 Training Development

Training development depends on the availability of various products and facilities (e.g., operations concepts, flight procedures, training simulators and mock-ups) which are delivered by various program organization units, based on requirements defined by the training organization.

Training organizations are responsible for the definition, development and implementation of the training programs and training schedules, and for the development of the training courseware comprising the training content, training tools and training materials (i.e., presentations, workbooks, manuals, computer-based training programs, etc.). Selection, training and

certification of training instructors are other important responsibilities of the training organizations.

All ISS partners agreed on a common concept for the development and implementation of training. **Instructional system development** (ISD) represents a very well-structured method of training development (see Figure 5.1.3).

At first, an analysis of all tasks to be conducted by the astronauts is made identifying the knowledge, skills and attitudes required from the astronauts to perform their activities.

During the design phase and based on the identified tasks, training objectives are defined. Based on their prerequisite requirements, the training objectives are grouped into lessons and courses, thus establishing a structured training flow. During this phase suitable training methods, required training tools and materials, and appropriate training evaluation concepts are identified.

Only then does the courseware development process start. Every phase of the training development ends with a formal review, so that the whole process is subject to continuous evaluation and deficiencies can be detected and corrected early in the development process. The training development concludes with a formal **training readiness review** (TRR) involving astronauts and training experts from all the ISS partners, who certify the lessons, training materials, facilities and instructors prior to the start of the implementation of the ISS program. The training implementation also includes a formalized feedback process in which the astronauts undergoing the training evaluate the quality of the instruction, training materials and facilities as well as subject matter expertise and the delivery skills of the instructors.

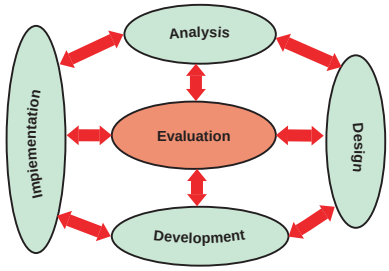


Figure 5.1.3: *Training development phases according to instructional system development.*

In addition, crew debriefings are conducted after every space flight in order to better link the training to the real operational needs as experienced by the astronauts during their mission.

5.1.2.3 Crew Qualifications

The training for all ISS systems is divided into three different qualification levels:

- **User Level:** Astronauts are familiarized with the major systems and learn everything that is relevant for safety aspects and for daily on-board routine.
- **Operator Level (includes the User Level):** This entails training for nominal and frequent operations as well as for off-nominal situations which require a rapid crew reaction.
- **Specialist Level (includes the Operator Level):** This covers training of infrequent nominal operations and training of appropriate crew reaction to off-nominal situations and malfunctions.

The training flow for the Columbus System Training (Figure 5.1.4) illustrates the allocation of lessons and courses within the three training qualification levels. The User Level training comprises cross-system introductory lessons on design and topology, operational concepts and monitoring and command capabilities as well as a system-specific course covering the relevant system knowledge and basic skills for all five Columbus systems (DMS: Data Management System, EPDS: Electrical Power Distribution System, TCS: Thermal Control System, ECLSS: Environmental Control and Life Support System, COMMS: Communications System). The Operator Level training deals with caution and warning responses and Columbus video system operations. Specialist training is dedicated to infrequent nominal operations like checkout and reconfiguration tasks as well as the reaction to system off-nominals.

5.1.2.4 Task-Oriented Astronaut Training

The minimum crew qualification levels and the main responsibilities which have to be available on the ISS during any time are predefined for every on-board system and spacecraft. For example, the Columbus systems require a minimum of one specialist, one operator and one user in a three-person ISS crew.

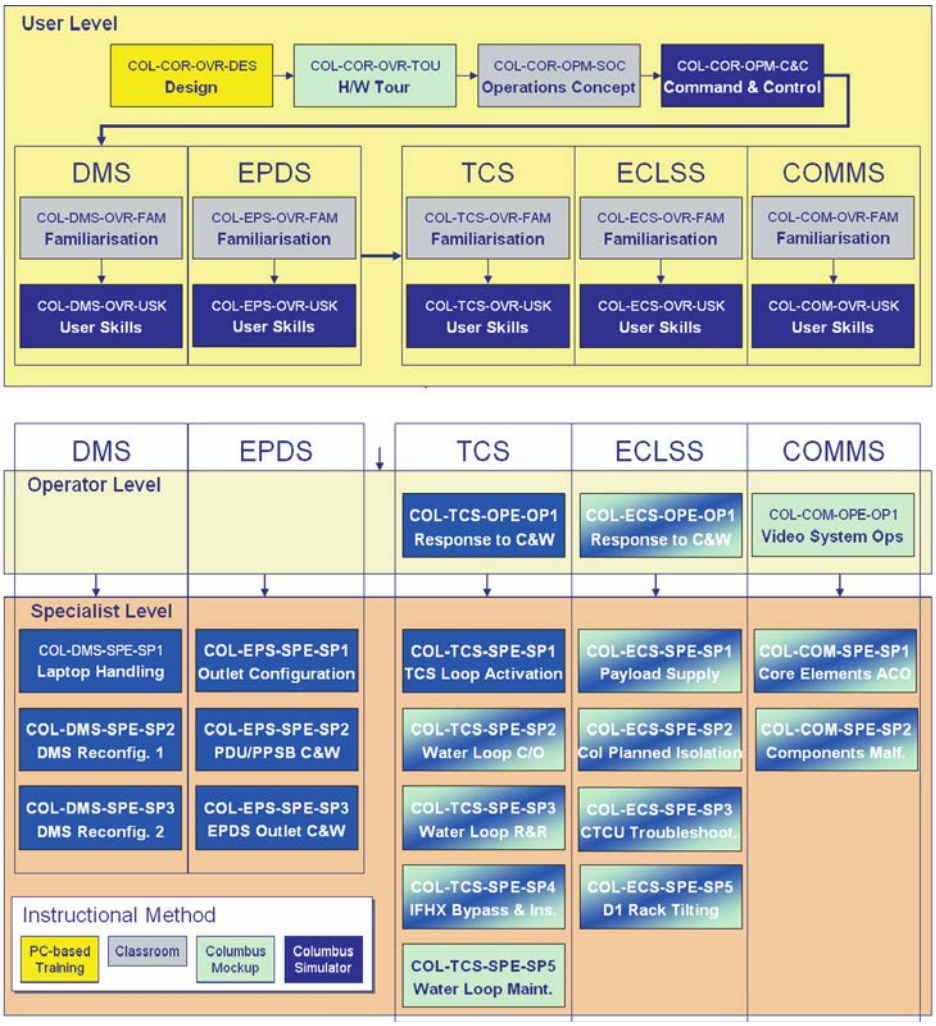


Figure 5.1.4: *Columbus System Training (Source: ESA).*

Based on these generic qualification requirements, the specific flight program for a mission and the prior experience and specific aptitudes of the assigned crew members, a CQRM is individually developed for every space flight mission before the training starts. The training of every crew member is individually tailored to match the distribution of tasks and qualifications as defined in the CQRM (see Figure 5.1.1).

5.1.2.5 Integrated ISS Training Plan

In order to achieve a high training efficiency, the ISS partners elaborate highly integrated ISS training plans, ensuring an adequate sequencing of training activities at the various partner sites and avoiding training redundancies and omissions. The implementation of this integrated training flow requires a high level of coordination and communication between all involved partners on an almost daily basis.

Figure 5.1.5 shows the integrated training flow for the ATV training, which is under ESA responsibility. Since ATV operation is supported by US and Russian systems, certain prerequisite training for ATV operations and rendezvous and docking is already conducted in the USA and Russia before the crew begins its first block of ATV rendezvous and docking training at the European Astronaut Centre in Cologne. Since ATV docks with the Russian service module,

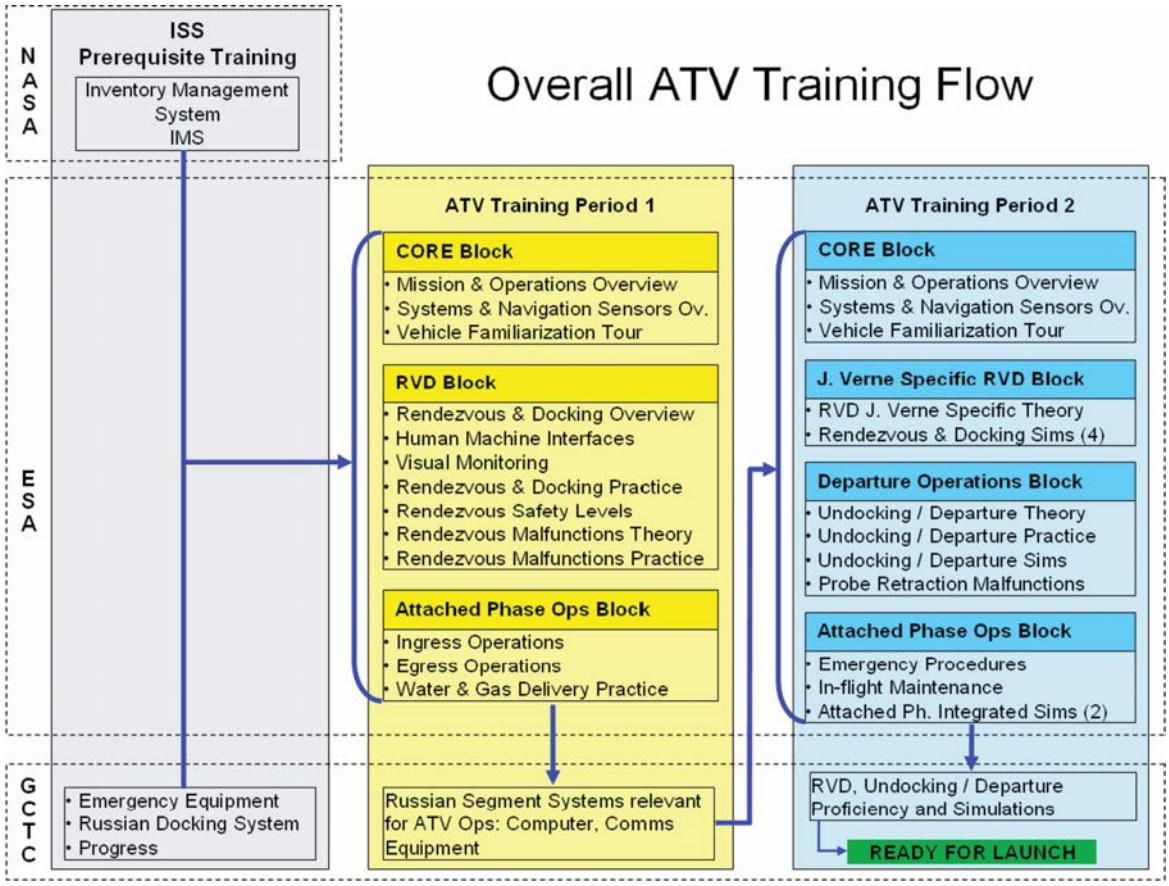


Figure 5.1.5: *The ATV Training Program (Source: ESA).*

some Russian system hardware is required during ATV-attached phase operations (e.g., to pump fluids and gases from the ATV's tanks into the station). This training, to be implemented in Russia, is a prerequisite for the second training block in Cologne.

After this block, repetition of some training units and final ATV simulations are implemented in Russia with cooperation between ESA and the Russian training center.

5.1.3 Infrastructure for Astronaut Training at the EAC

The training activities at EAC have two main target groups:

- Astronauts
- Ground support personnel.

In addition, all instructors and trainers have to be certified for the specific training programs and their duties as astronaut instructors.

EAC provides the infrastructure for all the different phases of astronaut training mentioned in Section 5.1.2.

5.1.3.1 General Purpose Training Infrastructure

The general purpose training infrastructure at EAC consists of classrooms equipped with state-of-the-art media technology.

EAC provides three classrooms for training. One of their key elements is the SmartBoard ®. These full-size

touch-sensitive back projections are fully integrated into the training lessons and presentations. Each of the different training phases also includes theoretical lessons which are conducted in the classrooms.

The following components of the general purpose training infrastructure constitute a fully integrated network and are maintained centrally:

- EAC Video Archive System (ViAS)
- EAC Voice Communication System (VoCS)
- EAC Backup System
- EAC GPS Timing System.

The EAC Video Archive System is composed of three main components. The video server unit, equipped with a 5 Tbyte hard-disk array, allows the digital recording of up to eight analog video signals. Simultaneously the video is streamed to the EAC LAN using data rates up to 4 Mbit/s. This allows observers to monitor the content in full-screen resolution on their personal computers. The sources for the eight streaming video channels and the corresponding audio channels are switched using a 64 × 64 (input × output) stereo audio/video matrix. Another 32 × 32 audio/video matrix is located in the training hall. This matrix allows preselection of the sources and destinations located in this area. Commercial off-the-shelf (COTS) software is used to record all eight channels into a circular buffer. The circular buffer allows approximately 60 days of recording before it starts to overwrite. Users have access to the recorded material via a Web browser interface. It allows viewing of all recorded material as in a video-on-demand system and provides the capability to generate MPEG2 clips from the recordings.

The EAC Voice Communication System is also a COTS system which is mainly used by air traffic controllers and media producers (e.g., broadcasters). This system handles up to 256 physical ports, which can be connected either point to point or as member of a conference. The user interface is provided either by fixed installations or via wireless mobile terminals. Thirty-two of the physical ports are connected to the core matrix of the Columbus Control Centre. This connection enables the ESA Ground Segment (GS) to have voice conferences during a simulation similar to the one used for real-time operations.

The EAC Backup System saves all the configurations of any computer-based training hardware at EAC to a centralized storage device. This technology assures the quick recovery of any device in case of a major incident.

The EAC GPS Timing System is used as the reference time base for all parts of the EAC training infrastructure. Using this reference a logged event in the training simulators, like the Columbus Trainer Europe (COL-TRE), can be correlated to an action monitored in the video recording of the astronaut training session.

5.1.3.2 Columbus System Training

The main components used for the Columbus system training are:

- Columbus Mock-Up (COL-MU)
- Columbus Trainer Europe (COL-TRE).

The COL-MU (Figure 5.1.6) is a high-fidelity copy of the orbital Columbus module. Each of the interfaces to be handled by the astronauts is physically available with the original "look and feel." All manual tasks for activation, deactivation, maintenance and replacement of parts are trained in the COL-MU. To meet the training requirements, each of the components, like connectors, filters, transport locks, etc., are available at least once in flight configuration.

The inner dimensions of the mechanical configuration (MECO) are equivalent to the flight module. All needed functional and data interfaces are available in the MECO. However, the concept of the COL-TRE is more complex (see Figure 5.1.8).

The core of the COL-TRE is a simulator consisting of the following components:

- Vital layer (original hardware of the Columbus flight system computers, DIFA) including the original flight software
- Simulation environment (functional models for temperatures, resources, cooling loops, etc.)
- Simulator control (control of the simulation, introduction of failure cases)
- Man–machine interface (MMI) computer
- Portable workstation (PWS)
- Payload simulators.

The vital layer is identical to the Columbus flight hardware and processor core for the original flight



Figure 5.1.6: *COL-MU (Source: EAC).*

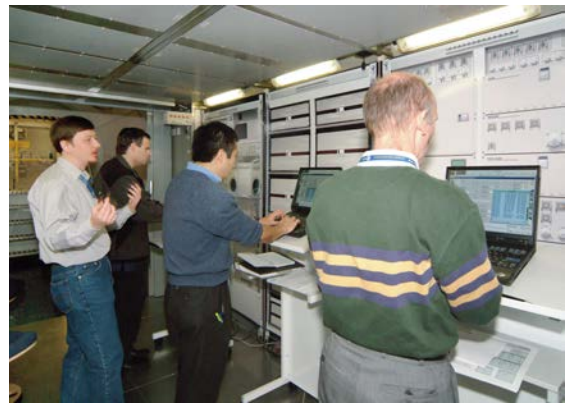


Figure 5.1.7: *COL MECO (Source: EAC).*

software. This system also manages the Columbus internal communication infrastructure, namely the MIL bus and the on-board LAN. The system communicates with the MMI system and the PWS. An astronaut manages and operates the Columbus system using the PWS as the main interface. The PWS laptop (Figure 5.1.9) runs the original on-board software with a graphical user interface to control and manage all functions of the Columbus system. The laptop is connected to the proper interfaces for power and data within the MECO. The original on-board software and the original vital layer software/hardware provide an authentic simulation environment which can easily be adapted to changes made within the software of the orbital unit.

Additional MMIs like switches for the lighting or video equipment are available in the MECO. But in contrast to the Columbus vital layer and PWS, these MMIs are provided by standard COTS equipment.

The simulation environment is linked between the vital layer and the external interfaces. This allows for manipulation of the data exchange between these

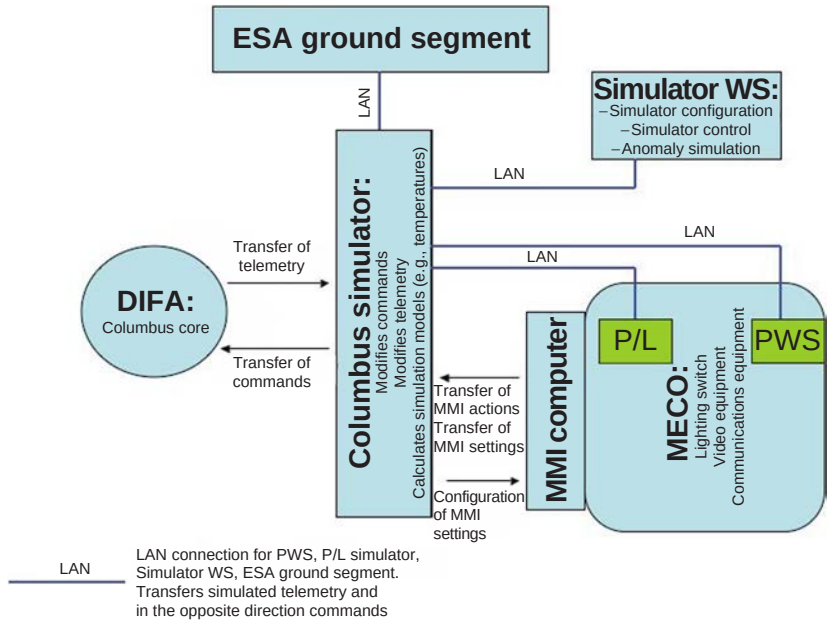


Figure *5.1.8: Schematic of the COL-TRE (Source: EAC).*

components for training purposes. Dedicated failure cases can be introduced into the system by tuning measures or actions. The simulation environment also contains software models for the environmental control system and system resources like power, cooling, gases and vacuum.

The COL-TRE generates the complete set of telemetry (TM) data (low- and medium-rate TM). An astronaut using the PWS can monitor nearly the complete TM. During so-called integrated simulations the COL-TRE is connected via ESA IGS (Interconnecting Ground Subnetwork) to the Columbus Control Centre in Oberpfaffenhofen, Germany.

5.1.3.3 Payload Training

The Columbus science laboratory hosts four European payloads and one passive rack:

- EPM European Physiological Module
- EDR European Drawer Rack
- Biolab Biological Laboratory
- FSL Fluid Science Laboratory
- ETC European Transport Carrier (passive).

Similar to the Columbus system training, the payload training concept foresees two different types of infrastructure for the different training purposes:

- "Stand-alone" payload trainer
- Payload simulators.

The "stand-alone" payload trainers are located in separate payload training booths. Comparable to the COL-MU, the "stand-alone" trainers represent a high-fidelity mock-up of the real flight hardware. Different from the Columbus system, the payload systems have a higher number of MMIs (switches, displays, connectors, etc.).

But also at the payload trainers all MMIs are available in high fidelity and their electrical and mechanical functions are identical to the MMIs used for the flight hardware. An instructor station is connected to each of the payload trainer racks to monitor the functions of the MMIs and to manipulate them for training purposes. A dedicated laptop serves as the main interface between the astronaut and the payload system. The laptops can be connected either to the stand-alone trainers or to the payload simulators and run a copy of the software used in the orbital systems.

Three payload simulators are integrated into the MECO of the COL-TRE. The simulators are software-based models of the payloads. A touchscreen device with back projection using a full Columbus 19-inch double rack provides the MMI between the user and the simulator. The payloads are graphically displayed in full size. Switches and all other user interfaces can be manipulated by using the touchscreen. The software models are fully integrated into the COL-TRE and to the ESA ground segment. During simulations the payload telemetry is forwarded via the COL-TRE to COL-CC or to the payload user centers, which are distributed across Europe. Accordingly, telecommands issued by any authorized facility are forwarded via the ESA ground station and COL-TRE to the payload simulators.

The payload simulators are used to train the nominal operations of the payloads as well as off-nominal conditions of the Columbus system as it relates to the connected payloads. The stand-alone payload trainers are mainly used to train manual tasks to be carried out by the astronauts, like activation, deactivation, exchange of probes and maintenance (see, e.g., Figure 5.1.10).

Simulations

The flight operations part of each space mission is handled by the ground operation teams. Each of the international partners in the ISS program is responsible for mission control and operations of its facilities in orbit. The different partners are also responsible



Figure *5.1.9: Portable workstation (Source: ESA).*

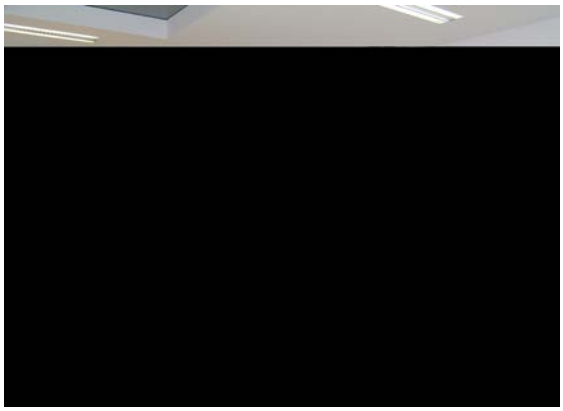


Figure *5.1.10: Biolab training model (Source: EAC).*

for the health of their astronauts on-board the ISS. The different facilities in Europe like the Columbus Control Centre (COL-CC), ATV Control Centre (ATV-CC), EAC and the payload centers share those responsibilities. Besides the training of astronauts, EAC is also responsible for the training and certification of the ground controllers and support personnel. Comparable to the astronaut training, the ground personnel have to go through a multilevel training and certification program. Only certified personnel are allowed to participate actively in mission operations tasks. An important part of the process is the participation in simulations. Three different simulations modes are distinguished:

- Stand-alone training und simulations
- Integrated simulations
- Joint integrated simulations.

The Training, Qualification and Verification System (TQVS) in COL-CC is used for stand-alone simulations. This system is nearly identical to the COL-TRE without the MECO and the Crew MMI. Stand-alone simulations are locally organized and managed at COL-CC for the training of the COL-CC ground controller team.

During integrated simulations the COL-TRE at EAC is used. The COL-TRE sends telemetry to the ESA ground station and receives commands as in the case of the Colombus orbital facility. The ground control teams are able to monitor the low- and medium-rate telemetry at their consoles and to uplink commands to the COL-TRE. This system does not produce high rate scientific data streams. On special occasions like system verification tests, a stream of static high-rate data is transmitted to the ESA ground station. Since the COL-CC distributed Voice Communication System (VoCS) is used in every simulation for all voice communications, the proper use of this system is an important part of the training and certification process. If required by the simulation/training objectives, a crew member or an appropriately trained crew member replacement is located in the COL mock-up or COL-TRE to perform the procedures and activities planned for the training session. This crew member will be monitored by the Colombus video cameras, whose images are linked to the ESA ground station. During all simulations, indications for the availability of communication links like acquisition of signal (AOS) and loss of signal (LOS) are simulated according to the original timeline of the simulated flight day.

During joint integrated simulations the ESA ground segment is connected to the NASA ground segment using the IGS. The Colombus Trainer US (COL-TRU) located at the Johnson Space Center in Houston, Texas, generates appropriate data streams. The COL-TRU is similar to the TQVS and COL-TRE but without any payload simulator capabilities. The COL-TRU is integrated into the NASA Space Station Training Facility (SSTF) and delivers its data directly to this facility. The main purpose of joint integrated simulations is to train all activities that need close cooperation between the different control centres, such as activation, deactivation and ISS emergency cases.

5.1.3.4 Automated Transfer Vehicle (ATV) Training

EAC hosts three facilities for ATV training:

- The ATV Mock-up (ATV-MU)
- The ATV Storage Trainer
- The ATV Crew Trainer (ACT).

The ATV Mock-up (Figure 5.1.11) consists of the ATV (pressurized module and propulsion module) and the Russian service module (Svesta). Both facilities are built to original scale. All cargo-handling and storage operations are trained in the ATV-MU or in



Figure *5.1.11: ATV mock-up (Source: EAC).*

the storage trainer. All ATV MMIs are available in high fidelity. The module is equipped with a complete set of cargo containers, the so-called cargo bags. The crew is trained in loading and unloading, fluid cargo handling (propulsion, gases, water) and emergency cases like fire and leakages. Only the ATV-relevant MMIs are available within the Russian service module, mainly the ATV crew workplace and the Russian communication equipment. All non-ATV-relevant crew tasks and crew interfaces in the Russian part of the ISS are trained in the Gagarin Cosmonaut Training Centre located close to Moscow.

The ACT is a complex simulator for all rendezvous and docking maneuvers of the ATV. Regularly the ATV approaches, docks and undocks in fully automated mode. But the on-board crew and the ground support crew in the ATV control centre in Toulouse, France, carefully monitor those phases of the ATV flight.

ATV transmits relevant telemetry data comprising attitude and control system information, important sensor data and video from the on-board camera. The crew is able to monitor the data and video at its ATV crew workplace. The crew is also able to select different telemetry parameters or different camera views on the displays. If the crew or ground personnel recognize that the vehicle is moving out of the limited corridor, both will be able to command an escape trajectory for the vehicle or even abort the complete approach maneuver.

The ATV crew workplace is installed in the Russian service module and also in the ATV training booth. By using an exact dynamic functional model and the original flight software of ATV, the ACT is able to generate all the telemetry data and computer-generated images which correspond to the video information. This makes the ACT and its MMIs an example of the use of virtual reality technologies in astronaut training.

In addition to the ACT facility, a software version of the ATV simulation is installed on a workstation and on the crew's on-board computers. A 3D viewer allows the trainee to have an overview of the positions and maneuvers of the ISS and the ATV.

Soyuz Trainer

Besides the Space Shuttle, the Russian Soyuz is currently the only available system providing crew transportation capabilities to the ISS. The main part of the Soyuz training is performed in Russia. The EAC Soyuz trainer provides capabilities for refresher training for the astronauts in case needed. The EAC Soyuz simulator is a software simulation. The crew MMIs comprise one original set of Soyuz steering devices and a touch-sensitive display like that installed in the latest version of the Soyuz capsule. Additional monitors show graphical simulations of the video images, the view through the Soyuz vehicle window, and a graphical simulation of the panel including all switches etc.

ISS System Training

The Columbus module is connected to Node 2 of the US segment of the ISS. All resources like power, cooling, data communication, etc., are provided through the interfaces between Columbus and Node 2. The main crew interface within the US part of the station is a laptop known as the Portable Crew Station (PCS). The American Segment Trainer (AST) is able to simulate the complete infrastructure including Windows server architecture. Especially during the installation of the Columbus module on the station and during the first activation, the crew uses the PCS as a main interface. Furthermore, the AST includes the ISS Inventory Management System (IMS), which is also used for the management of items delivered by ATV and used during the ATV cargo training.

Neutral Buoyancy Facility (NBF)

Europe's largest neutral buoyancy facility is located at EAC. The dimensions of the tank are: 22 m length,



Figure *5.1.12: NBF and SSDS system (Source: EAC).*

17 m width and 10 m depth. The facility includes a completely equipped control room with appropriate infrastructure.

Initially planned for training in EVA using a European crewed vehicle, the facility is now mainly used for basic training of the astronauts. The objective is still to prepare them for EVA, but the vehicle and EVA-suit-specific training is provided by either the Russians or the Americans. The basic EVA training consists of SCUBA training and EVA fundamentals training. The EVA fundamentals training is performed using a full-face mask system (SSDS). The breathing air is supplied via a harness that also allows bidirectional communication between the EAC voice system, the control room and the trainee (Figure 5.1.12).

A number of fixed and mobile cameras allow the complete supervision and documentation of all activities in the pool. The NBF video and audio systems are integrated into the EAC VoCS and EAC ViAS.

Bibliography

[5.1.1] US Air Force Instruction Manual *AFMAN 36-2234*, Instructional System Development (ISD), 1993.

[5.1.2] Station Program Implementation Plan (SPIP), Vol. 7. *Training, rev. A*, 1997.

[5.1.3] Multilateral Training Management Plan (MTMP), Vol. 1. *SSP 41184-01*, 1998.

[5.1.4] Bolender, H., Bessone, L., Schoen, A. *et al.* EAC Trains its First International Astronaut Class. *ESA Bull.*, **112**, 50–55, November 2002.

[5.1.5] Bolender, H., Bessone, L., Schoen, A. *et al.* Training ISS: EAC Training its First International Astronaut Class. *On Station*, No. 11, 12–13, December 2002.

[5.1.6] Bolender, H., Clervoy, J.-F., Schoen, A. ATV Ahoy! – Crew Training for the First ATV. *On Station*, No. 86, 15–17, November 2004.

[5.1.7] Damian, K., Tognini, M. The European Astronaut Centre: A Look Back at 15 Years of Astronautical Expertise. *ESA Bull.*, **123**, 55–60, 2005.

[5.1.8] Eichler, P., Seine, R., Khanina, E. *et al.* Astronaut Training for the European Contributions Columbus Module and ATV. *Acta Astronaut.*, **59**, 1146–1152, 2006.

[5.1.9] Bolender, H., Stevenin, H., Bessone, L. *et al.* Preparing for Space – EVA Training at the European Astronaut Centre. *ESA Bull.*, **128**, 32–40, November 2006.

5.2 Life Support Systems

Willigert Raatschen

5.2.1 Functions of a Life Support System

5.2.1.1 Introduction

The life support system of a habitat, whether a space station, a space vehicle or a base on a planet, has the purpose of providing an environment and resources which preserve human life in health and safety for the duration of a mission. On Earth, the **biosphere** is the human life support system. Earth's biosphere is an almost closed system as to mass transfer, but an open system as far as energy exchange with space is concerned [5.2.2]. It is the task of a life support system to assure the biological autonomy of humans in an isolated environment, while creating

a physiologically acceptable environment which makes residence there comfortable. This includes attention to vibration and noise levels, protection from ionizing and nonionizing radiation, electromagnetism and gravity or the effects of microgravity. Thermal comfort, adequate air quality and safe handling of all equipment in the habitat itself are likewise critical.

Of course, there are different understandings of which functions belong to a life support system. For the Russians the life support system includes, for example, the storage and preparation of food, refrigerators and freezers, extravehicular activities, personal hygiene and housekeeping functions. For NASA these areas are covered under "crew systems." There are also differences in how individual subfunctions of the life support system are understood, a topic which will not be addressed here.

Figure 5.2.1 summarizes the classic functions of a life support system and lists on the right related functions which do not typically belong to a life support system but which will also be discussed for the sake of completeness.

The following discussion describes the life support system of the International Space System (ISS), as the only continuously occupied space habitat at present. The individual functions can be applied to space vehicles as well as lunar or Martian habitats, where, however, the technical realization can be very different. There can also be large differences in how critical individual functions are considered to be. For example, for extravehicular activities on the ISS, bringing in contamination plays a secondary role, but the input of dust and possibly bacteriological or biological trace elements on Mars is a very important concern which accordingly has a significant influence on the technical implementation of the life support system.

5.2.1.2 The Atmosphere

Assuring an adequate atmosphere is the task of the **atmosphere control and supply system** (ACS), which must fulfill the following requirements:

1. Total pressure monitoring
2. Oxygen (O_2) partial pressure monitoring
3. Storage and supply of nitrogen (N_2) and O_2
4. Pressure equalization between neighboring modules
5. Detection of sudden pressure loss and initiation of countermeasures.

Table 5.2.1 summarizes the partial and total pressure levels on the ISS.

5.2.1.3 Pressure Control, Relief and Valves

Using the example of the European **Columbus module** on the ISS, Figure 5.2.2 shows the arrangement of valves on the external shell. Two **overpressure valves**

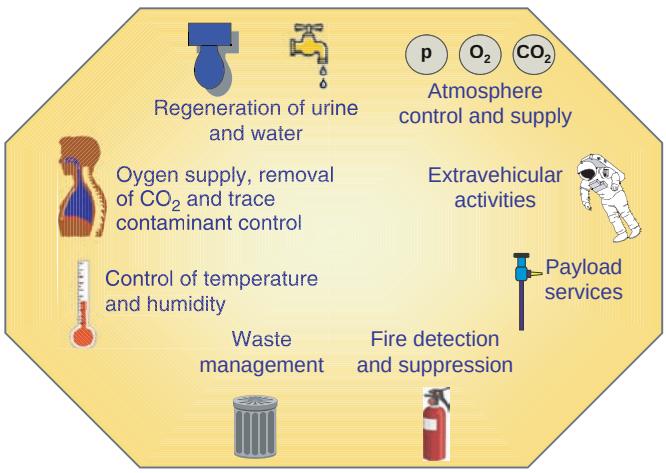


Figure 5.2.1: *Classic tasks of a life support system; related functions at right (Source: ASTRIUM).*

Table 5.2.1: *Air pressure requirements for the ISS.*

Parameter	Requirements
Total pressure	97.9 to 102.7 kPa 95.8 kPa minimum
O_2 partial pressure	19.5 to 23.1 1 kPa
N_2 partial pressure	< 80 kPa
Maximum inside–outside pressure differences; exceeding permissible limits triggers opening of the positive pressure relief valve	103.4 kPa

(**positive pressure relief valves**) automatically blow air from the module to the outside as soon as an inside pressure of 103.4 kPa is exceeded. Each overpressure valve consists of two valves connected in series. The outer valve is always open and can be either electrically or manually closed if the inner valve leaks or is damaged.

Of the four **underpressure valves** (**negative pressure relief valves**) on Columbus, two are installed on the closed front end and the other two in the fresh air and return air duct, called the intermodular ventilation (IMV) ducts. The latter are exchanged for IMV valves when Columbus docks with the ISS.

The underpressure valves are required should the outside pressure exceed the inside pressure by more than 26 to 33 hPa, for example when Columbus is transported from Bremen, Germany, to the Kennedy Space Center, Florida, or in case a Space Shuttle launch is aborted, followed by an emergency landing.

The underpressure valves function mechanically. In order to avoid later leakage problems, a sealing cap is screwed on after the shuttle docks with the space station.

In order to rapidly evacuate the module in case of fire, two redundant valves located in the external shell can be opened electrically (cabin depress valves). The outside duct is heated to prevent ice formation when warm, moist air is released.

A **pressure equalization valve** is integrated in the door to the neighboring ISS node. It is operated manually in order to achieve pressure equalization after docking and before opening the hatch. It can also be used to extract air samples for analysis. The **vent line dump assembly** has the purpose of drawing off process gases from the experiments and releasing them into space. Additional vent lines are needed to draw off the carbon dioxide which has been filtered out of the internal air and the hydrogen which is generated during **electrolysis**. The outlets of the vent lines have to be designed in a way that avoids their exerting any additional impulse (plume impingement) on the space station during the venting operation. In most cases the terminal nozzle is located far enough from the module shell to avoid deposition of waste gas residue on the shell. Space vacuum is provided for experiments from the vacuum ducts and associated vacuum line assembly

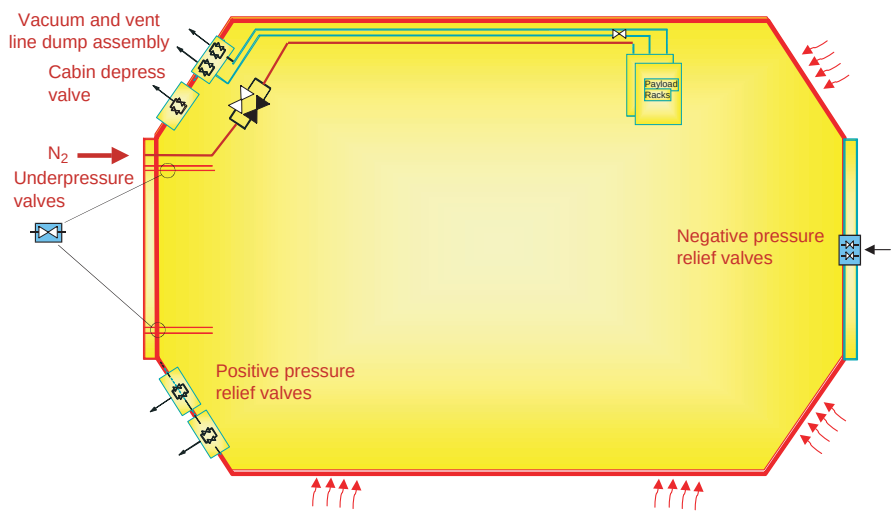


Figure 5.2.2: *Valves on the external shell of the Columbus ISS module, including the valves to vent gases from the experiments (Source: ASTRIUM).*

valves. The waste gas and **vacuum valves** are structurally similar to the cabin depress valves, in that the former are not heated. All electrically operated valve pairs are supplied and controlled by different electric circuits and control units for reasons of redundancy.

5.2.1.4 Monitoring Air Composition

The air in the ISS has the same composition as on Earth, about 79% nitrogen (N_2) and 21% oxygen (O_2). The oxygen is inhaled by the astronauts and eventually animals on-board and must be resupplied, see also Section 5.2.1.6. Air can also escape because of leakage, reducing the total pressure in the station. Oxygen is resupplied by water electrolysis; the amount is based on measurement of O_2 partial pressure. The amount of nitrogen to be resupplied is determined by a measurement of total pressure. It is stored in a gaseous form in pressurized tanks.

In the Columbus module **oxygen concentration** is measured with an **oxygen sensor** located in the incoming air duct. In addition, from each module on the US side of the station there is a **trace gas sample line** to the US laboratory (Figure 5.2.17 below shows a schematic of all the ISS modules). There the air is analyzed with a **mass spectrometer** (multiconstituent analyzer, MCA) as to the concentrations of nitrogen (N_2), oxygen (O_2), hydrogen (H_2), carbon dioxide (CO_2), methane (CH_4) and water vapor (H_2O). Measurement of N_2 partial pressure immediately reveals any leakage in the outer shell.

5.2.1.5 Ventilation and Air-conditioning

Since there is no natural air convection under microgravity conditions, the air in a space station would be immobile and the astronauts would breathe in most of the CO_2-laden air which they had just exhaled. For this reason, air-conditioning is not only a comfortable convenience but absolutely vital, and the air circulation ventilator is accordingly redundantly designed, see Figure 5.2.3. The limiting factor for the extent of ventilation in the module is the thermal **comfort range** for humans, which is primarily determined by the air temperature and velocity, and the humidity. The indoor climate requirements are listed in Table 5.2.2. The odd numbers for air temperature and velocity are a consequence of conversion from US units to SI units.

Figure 5.2.3 shows the **ventilation concept** for Columbus. Via the fresh air inlet (IMV supply air) the air processed centrally in the US laboratory is supplied to the module through the IMV incoming air ventilator. The 210 m^3/h volume flow is designed to supply three astronauts in the Columbus module. But in order to generate an air velocity that is high enough there, the same volume must also be recirculated via the air **recirculating fan**.

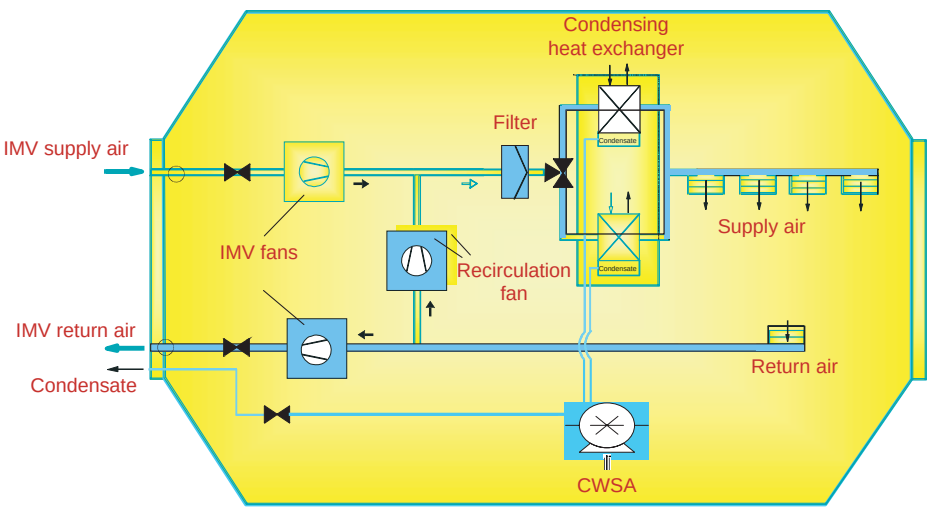


Figure 5.2.3: *Schematic of the air distribution and air-conditioning system for the Columbus module (Source: ASTRIUM).*

Table 5.2.2: *Columbus air requirements.*

Air temperature	18.3 to 26.7 °C, Adjustable by the astronaut
Humidity	25 to 75% r.h.
Air velocity	0.076 to 0.203 m/s in 67% of the module volume 0.036 to 1.016 m/s at all locations

Before air enters the module it passes through a **particle filter** (HEPA), after which it is cooled in a **heat exchanger module** and dried, if necessary. It is then uniformly distributed to the inhabited areas via two sets of four **ceiling outlets (diffusors)**. Used air is sucked out of the module through an air vent in the front floor area, whereby 50% returns to the cabin via the recirculating air fan and 50% is routed to the US laboratory for processing, see Figure 5.2.4.

Since heat release via the outer shell is very limited, and also considerable waste heat is generated in the cabin by the equipment and the astronauts themselves, there is more often a need for **air-conditioning** than for heating the cabin. The dominant **humidity sources** on a space station come from astronauts, who exhale moisture and perspire, and from moist cloth or **laundry** drying in the cabin air. In order not to exceed a maximum relative humidity of 75%, the cabin air has to be dried. This is accomplished by a **condensing heat exchanger** which both cools and dries the air.

The upper limit for relative humidity is chosen so that no condensate can form on the inner surfaces of the outer shell. Damp substances and condensate would lead to **microbial growth** and **mold**, which was a major problem on the MIR space station. The lower relative humidity level of 25% was chosen to minimize the risk of electrostatic charge and subsequent discharge sparking caused by air that is too dry, and to keep the mucous membranes of the astronauts from drying out.

The condensing heat exchanger on Columbus, see Figure 5.2.5, is an air–water heat exchanger and consists of two identical transfer blocks. Cold water from the low-temperature water circuit (low-temperature

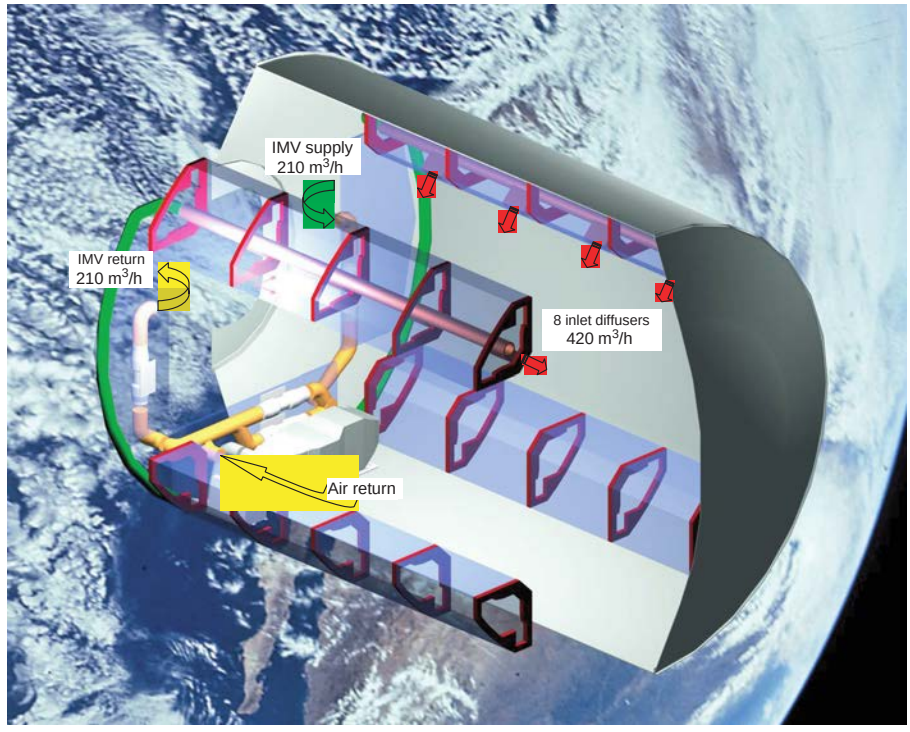


Figure 5.2.4: *Air duct layout for the Columbus module on the ISS showing the air outlets (Source: ASTRIUM).*

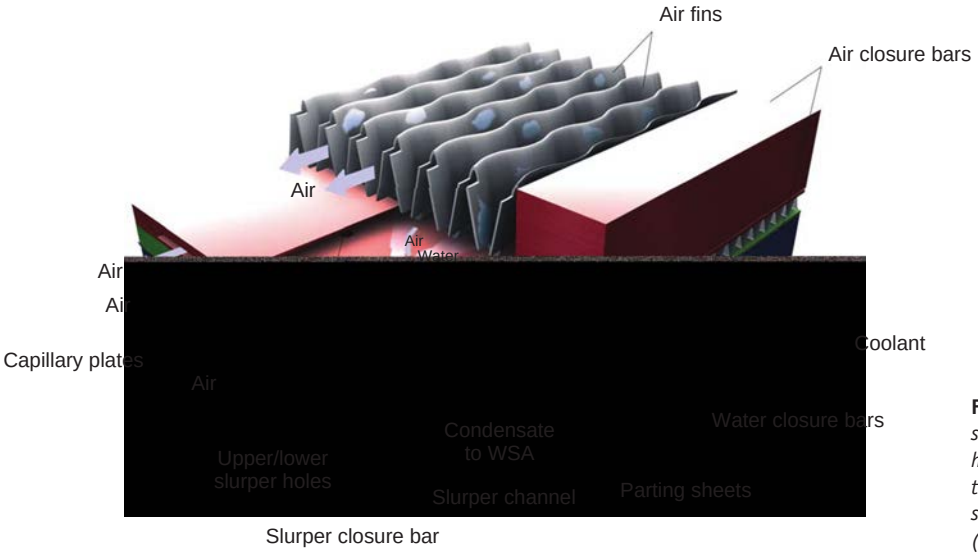


Figure 5.2.5: *Design schematic of the Columbus heat exchanger with selective removal of condensate via the slurper section (Source: ASTRIUM).*

loop, LTL) of the ISS flows in on the water side and cools down the cabin air. A **temperature control valve** (TCV) is nothing but an air flap which divides the incoming air stream, directing it to the two transfer blocks. **Cooling water** flows through only one of the two transfer blocks. If the need for cooling increases, then the TCV directs more air through the cooled block, and vice versa. The cabin air cannot be actively heated. The **heat loads** are usually sufficient to increase the air temperature.

When moist air is cooled, water drops usually form on the cold surfaces of a heat exchanger. Under microgravity they would be swept along by the air stream and return to the cabin via the air outlets, which could lead to short circuits in the electrical equipment. It is therefore critical that any **condensate** forming on the heat exchanger be safely removed. The technology applied for Columbus is shown in Figure 5.2.5.

The air fins on which the droplets condense have a biocidal, hydrophilic coating. Less than 1 μm thick, this hydrophilic layer causes the water droplets to flow over the fin surface as a film. This liquid film is then transported via the air stream to the end of the fin, where it is sucked between the capillary plates due to **capillary forces** and drawn off through holes as a condensate–air mixture with the help of the **condensate water separator assembly** (CWSA). A **liquid carry-over sensor** (LCOS) downstream of the heat exchanger reacts whenever water droplets are in the air stream.

In order to prevent microbial growth on the moist surface of the heat exchanger, the coating contains silver oxide as a biocidal substance. During its 10-year lifetime it is slowly released to the condensate, thereby reliably preventing the growth of microorganisms. In addition, at intervals of three to four weeks the cooling is switched from one to the other heat transfer block. This gives the moist block a chance to dry out, which also hinders the growth of microorganisms.

The CWSA in Figure 5.2.3 is a **centrifugal separator** which conveys the condensate back to the US laboratory for processing. Whereas the modules on the US side of the ISS (US laboratory, **Nodes 2 and 3**, **JEM**, **Columbus**) all have their own condensing heat exchangers, the air in the Russian modules (**FGB1**, **SM**, **MLM**) is cooled and dried centrally in the service module (SM).

5.2.1.6 Air Revitalization and Trace Contaminant Removal

Processing air to be breathed, filtering and removing gaseous pollutants, monitoring the composition of the air, and emergency care in case a critical component fails are the functions of the **atmosphere revitalization system**.

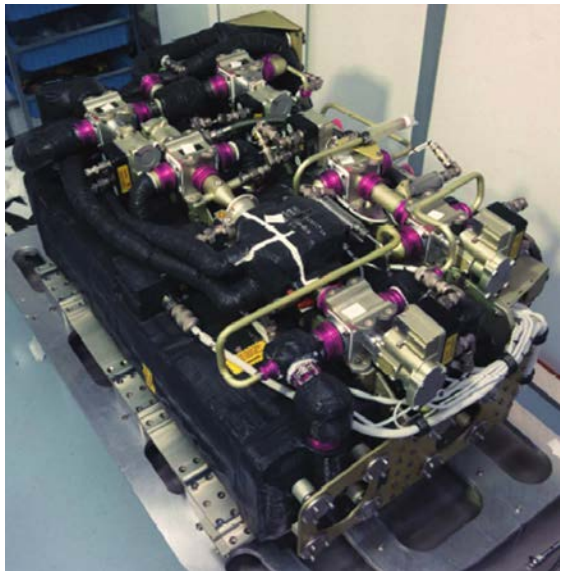


Figure 5.2.6: *The US carbon dioxide removal assembly (Source: HONEYWELL).*

Carbon Dioxide Removal

An astronaut exhales on average almost exactly 1 kg of carbon dioxide per day. This means that if six astronauts are on the ISS, then 6 kg per day, or 2190 kg of CO_2 per year, have to be filtered out of the cabin air and disposed of. This is the function of the carbon dioxide removal assembly (**CDRA**) in the US laboratory and of **VOSDUKH** on the Russian side; see Figure 5.2.6. Both systems make use of the adsorption capability of **zeoliths** to extract CO_2 from the air. Each system has two beds. While bed 1 adsorbs, bed 2 desorbs, and vice versa. Before adsorption the cabin air has to be dried. **Desorption** is accomplished by heating and discarding the CO_2 into space.

On the European side, another technology is being developed in the closed loop **air revitalization system** (ARES). It is based on the capability of carbon dioxide to adsorption on the surface of **ASTRINE™**, a specially developed ion exchange resin. As in the case of zeoliths, cycling takes place between at least two beds; that is, one bed is always in the adsorption phase while the other desorbs. The air does not have to be dried or processed in advance when using solid amines because the bonding is only slightly influenced by humidity.

Regeneration of a bed saturated with CO_2 is accomplished with slightly overheated water vapor (105–110 °C). The water vapor flows over the ASTRINE™, acting somewhat like a piston which pushes the desorbed CO_2 in front of it. The advantage of **water vapor desorption** is that it can take place at ambient pressure (approx. 1 bar) and the desorbed CO_2 can be easily used to close the **oxygen cycle** (see the section on oxygen reclamation below). If a vacuum is required for desorption, then the CO_2 is released to the outside and any chance for reutilization is lost, or additional devices like vacuum pumps and CO_2 compressors are required for its reuse.

Oxygen Supply

The supply of **breathing oxygen** on the ISS is provided chiefly via water **electrolysis**. The primary backup system consists of a **solid fuel oxygen generator** (SFOG); there is a second backup system on the US side consisting of two pressurized oxygen tanks outside of the US airlock.

Oxygen is also required for medical care and for the emergency breathing masks, for which there are supply connections in every module. The masks are worn in emergencies when pollutants are released, such as from fires. The oxygen from the pressurized tanks is primarily used to prepare the astronauts for extravehicular activities, see Section 5.2.1.9.

In water electrolysis, liquid H_2O is broken down into its molecular components, hydrogen H_2 and oxygen O_2, with the help of electricity:

$$2\,H_2O \rightarrow 2\,H_2 + O_2$$

Water electrolysis can be carried out using various technical processes, whereby there are three processes qualified for space applications. The Russian Space Agency (RSA) uses **alkaline electrolysis** with a potassium hydroxide solution and a circulation pump system, as was done on the MIR space station. NASA has an electrolyzer [5.2.3] based on **proton exchange membrane** (PEM) technology, while ESA uses **fixed alkaline electrolysis** (FAE) [5.2.4]. Figure 5.2.7 shows photographs of the various electrolyzers.

Oxygen Reclamation

In an open loop, as was the case in the ISS when this book went to press, the desorbed carbon dioxide and

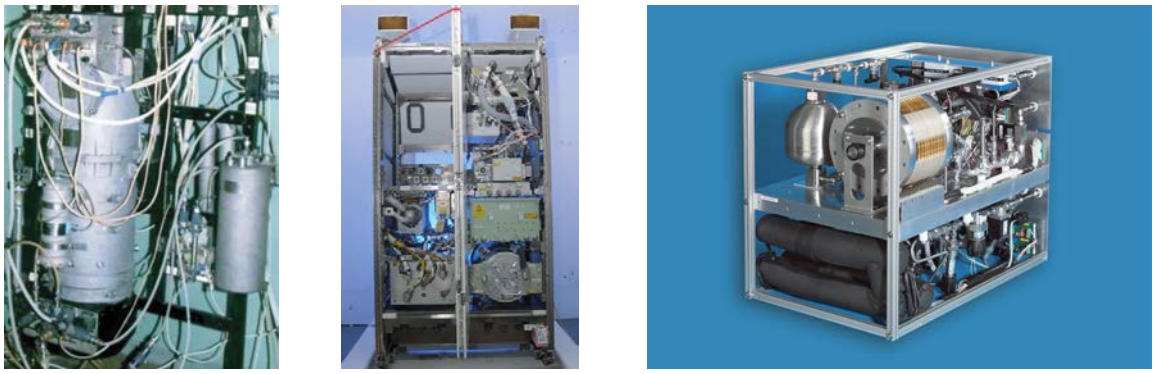


Figure 5.2.7: *The Russian electrolyzer ELEKTRON (left), the US electrolyzer (center) and the European electrolyzer (right) (Sources: Niichimmash, Hamilton Sundstrand Space Systems, and ASTRIUM, respectively).*

the hydrogen by-product from the electrolysis process are vented to space. With the help of the **Sabatier reaction** most of both substances could be utilized, because the material loop could then be closed; see section 5.2.2.

Figure 5.2.8 shows the incorporation of the Sabatier process in the life support system. The carbon dioxide and the hydrogen are fed through a catalyst bed at temperatures between 300 and 500 °C, producing water vapor and methane, CH_4:

$$CO_2 + 4\,H_2 \rightarrow CH_4 + 2\,H_2O$$

This is an exothermic reaction initiated merely by the addition of energy. The water which is generated can be used in the electrolyzer; there is no use for the methane on the ISS so it is discarded into space. The conversion efficiency for the hydrogen is between 90 and 100%, depending on the operating point. The reactor (Figure 5.2.9) is run under subpressure to prevent any hydrogen leakage into the cabin.

Methane Pyrolysis

If one wants to avoid having to bring additional hydrogen on-board the space station, then only about 58% of the carbon dioxide can be converted in the Sabatier reactor using electrolysis and hydrogen; see also Section 5.2.2. The oxygen and hydrogen cycles can be completely closed by adding a subsequent step involving the **pyrolysis** of methane. The methane is split into its components of carbon and hydrogen at high temperatures with the addition of energy. This hydrogen is then fed into the Sabatier reactor; the amount is adequate to convert the entire CO_2 flow. The entire process is illustrated in Figure 5.2.10.

Monitoring and Removal of Particles and Gaseous Pollutants

Before air enters the condensing heat exchanger, particles are filtered out with the help of a **high-efficiency particle filter** (HEPA, Class EU6).

In a closed habitat, humans emit **gaseous pollutants** (defined as gases other than O_2, CO_2 and water vapor) to the ambient air; materials outgas; processes (cleaning, gluing, soldering, cooking, etc.) emit gases, as can leaks, for example from the cooling system. Reaction products from the attitude control jets can be transported into the module via the space suits as a consequence of extravehicular activities.

The theoretical and actual pollutant cocktail on a space station can be an exotic mixture. In the procedure used on the ISS in the US On-Orbit Segment (USOS), shown in Figure 5.2.11, gaseous pollutants with high molecular weights are adsorbed on activated charcoal. Contaminants with low molecular weight such as CO, CH_4, H_2 and other hydrocarbons are combusted to water and carbon dioxide via a **high-temperature catalytic converter**. Since there is a risk that, for example, Freon refrigerant from the Russian cooling cycle decomposes into toxic by-products because of the high temperatures in the catalytic converter, a **lithium hydroxide bed** is

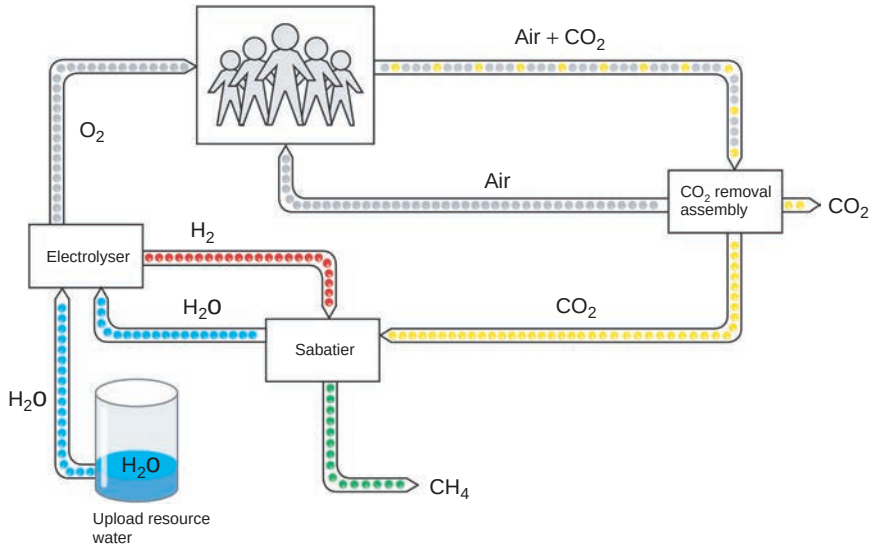
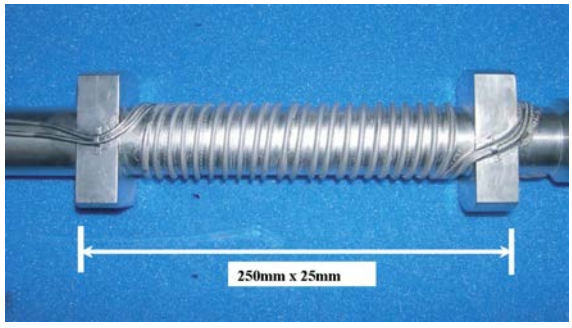


Figure 5.2.8: *Schematic of oxygen reclamation using a Sabatier reactor (Source: ASTRIUM).*



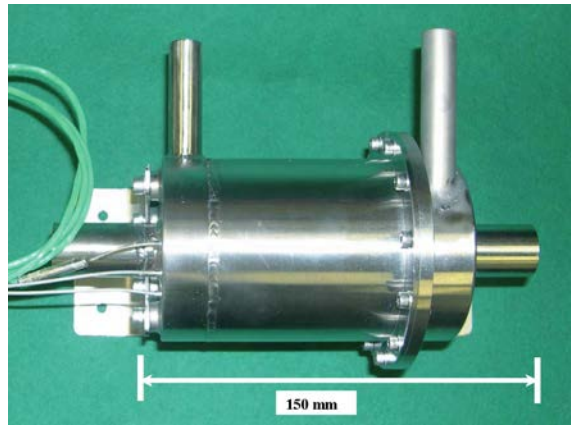


Figure 5.2.9: *The Sabatier reactor as a component of ARES (Source: ASTRIUM).*

connected downstream of the high-temperature catalytic converter.

Particle filters as well as **activated carbon and LiOH cartridges** are replaced at regular intervals as consumables. A **mass spectrometer** serves to monitor gaseous pollutants in the US laboratory. Every few minutes the air in each module is analyzed for O_2, N_2, CO_2, H_2, CH_4 and H_2O content making use of sampling ducts leading to the Columbus module, to JEM and to Nodes 2 and 3. In special cases measurements can be made at intervals of 2 seconds in one module. All other volatile **organic substances** are detected and identified with the volatile organic analyzer (VOA), which consists of a **gas chromatograph separator** and an ion mobility spectrometer (IMS).

5.2.1.7 Water and Urine Processing

Dehumidifying the cabin air involves the removal of about 1.5 l of water and an additional 1.3 l of urine per day and astronaut. The processing of this amount of water is one of the most important functions of a life support system. The components of the **water management system** are:

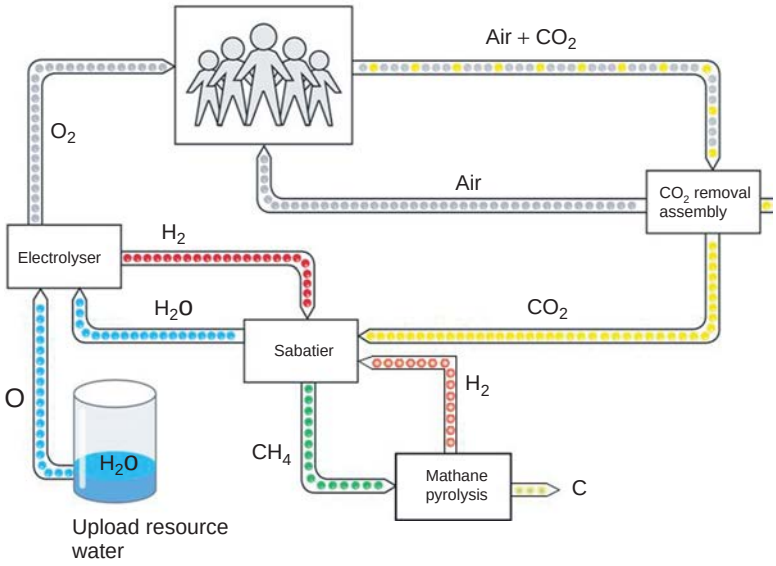


Figure 5.2.10: *Closed hydrogen and oxygen cycles in a human habitat (Source: ASTRIUM).*

1. A waste water drainage system, which is needed to dispose of excess water into space in case the water processing equipment is defective (or has not yet been installed during the habitat construction phase).
2. Water tanks to store condensate from the dehumidifier, water for washing, fresh water and water from the fuel cells if a supply vehicle uses this energy source (as is the case for the Space Shuttle) and deposits the generated water in the space station or habitat.
3. A water distribution network, which usually consists of three separate components:
   (a) A condensate water network in which all water from the various habitat areas or modules is collected for centralized water processing.
   (b) A waste water network to collect washing water (oral hygiene, personal hygiene, wet shaves), waste water from the space suits, animal experiments and urine.
   (c) A freshwater network for the drinking and process water supply (EVA suits, electrolyzer).
4. Water processing.
5. Urine processing.
6. Freshwater quality control.

Water Tanks

The water tanks are usually **blow tanks** in an aluminum container or have a flexible external shell. The tank is emptied by blowing in air or nitrogen.

Water Processing

The processing of waste water from condensate and washing water begins with its degassing, after which particles are removed with a filter. The water is then routed over sorption and **ion exchange beds** to remove pollutants. The next step is to heat the water and pass it over a catalytic converter in order to oxidize organic components of low molecular weight and organic acids, which requires the addition of oxygen. After cooling, the water is again routed over an ion exchange bed to remove the by-products of the oxidation process. Excess oxygen is removed at the following membrane stage, after which iodine or silver oxide is added as a preservative.

Disposal and Processing of Sewage

Urine is first preprocessed at room temperature and then treated with water purification equipment. The preprocessing consists of the addition of ozone and sulfuric acid to suppress ammonia formation during

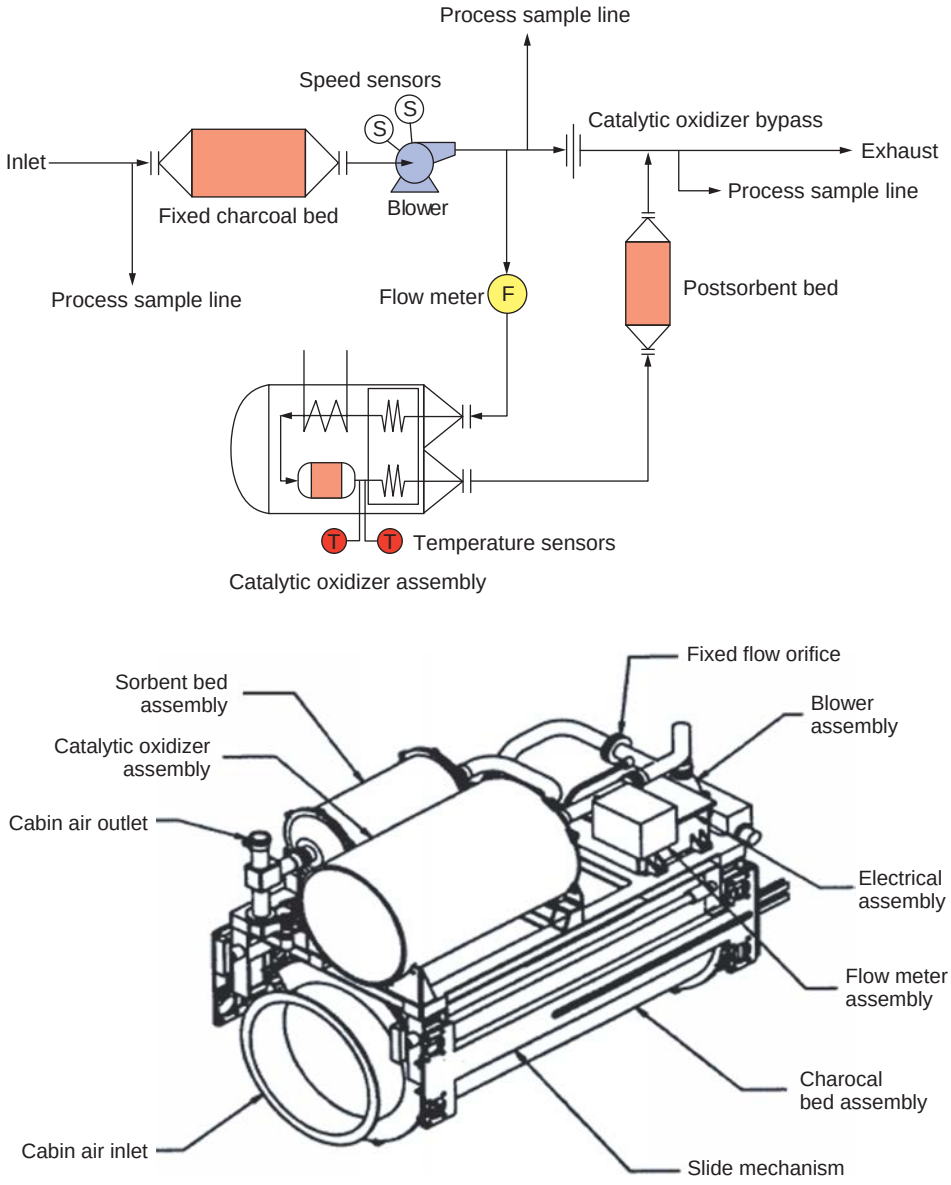


Figure 5.2.11: *Trace contaminant removal in USOS on the ISS: simplified process diagram (above) and flight hardware configuration (below) (Source: J.L. Perry).*

storage. For ease of handling, the ozone and sulfuric acid are usually in tablet form and installed with a filter as part of a hose unit. The urine is then vaporized, concentrated and condensed via a downstream distillation column in order to obtain pure water. This product is then treated in the water purification equipment.

Figure 5.2.12 shows the basic design of a **space toilet**, which has to function in microgravity. The upper schematic shows a compactor for the bowel movement. When the toilet lid is raised, a ventilator turns on which draws air into the compactor to prevent odors from escaping. Fecal matter is collected in a plastic bag usually sufficient for 25 to 30 defecations. When it is full it is closed with a clip and can be disposed of. The air that has been sucked in is passed over an odor filter and returned to the cabin. The fans are usually available redundantly.

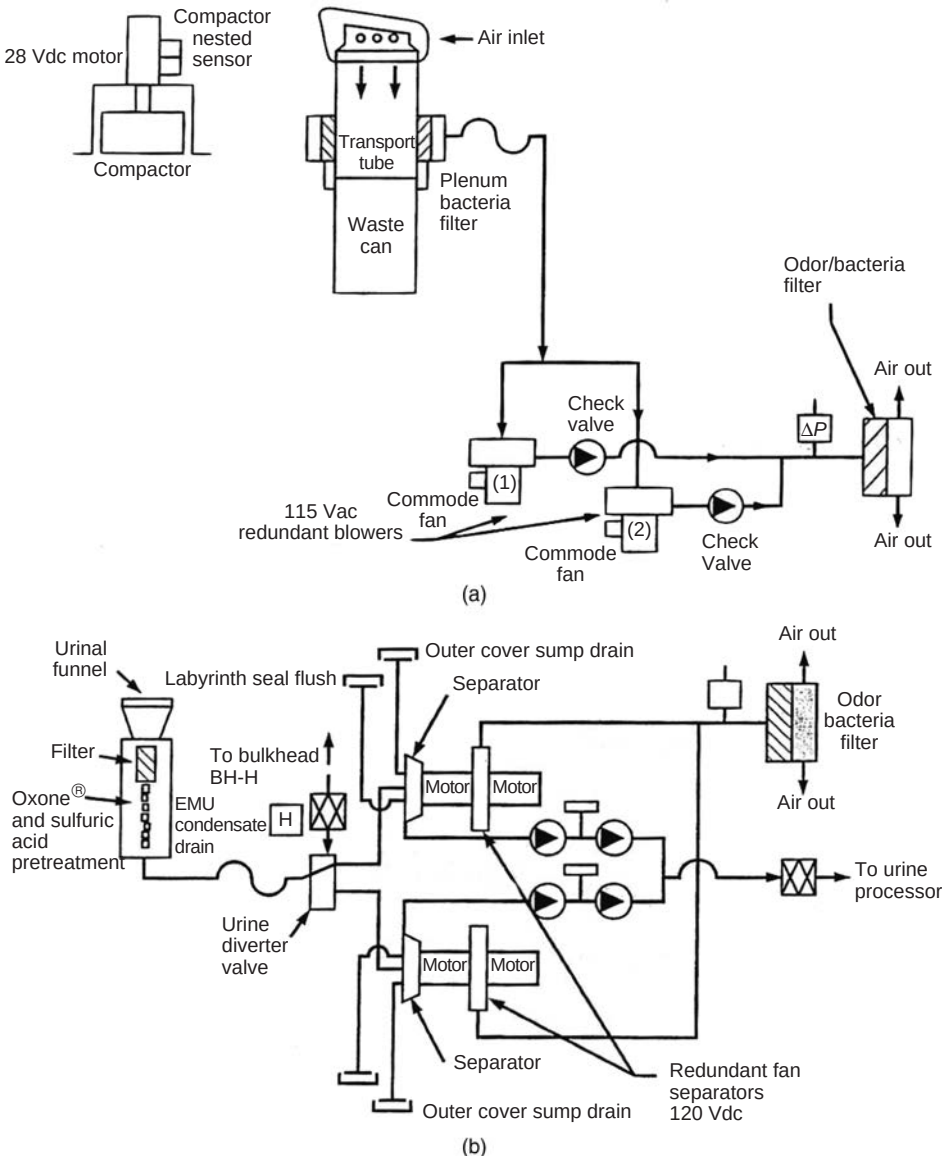


Figure 5.2.12: *Schematic of a space toilet, facilitating bowel movements (top) and collecting urine (bottom) (Source: P.O. Wieland).*

The lower schematic shows how the urinal functions. Urination is usually into a hose. Here, too, air is drawn into the hose to draw off any odors or droplets. For hygienic reasons each astronaut has his or her own adapter. After the urine flows over the ozone and sulfuric acid tablets the urine–air mixture is separated. The air flows over the **bacteria and odor filter** and then back into the cabin, and the urine is routed to the urine processor. Photographs of the Russian toilet on the ISS and the US toilet on the Space Shuttle are shown in Figure 5.2.13.

Water Processing on the ISS

On the Russian side a distinction is made between **drinking water** and **technical water**. Drinking water is exclusively obtained from the dehumidifier condensate and from washing water, has minerals added and is used only for drinking. The technical water is



Figure 5.2.13: *Russian toilet in the ISS service module (left); US toilet in the Space Shuttle (right).*

supplied from processed urine and is used only for electrolysis and flushing toilets.

On the US side (not yet in operation as this handbook went to press), condensate, washing water and the preprocessed urine are cleansed in the same water processing unit and the product has drinking water quality, whether it is used for drinking, washing or in the electrolyzer. The only difference is that the water for drinking has mineral additives.

Whereas the water on the Russian side is preserved with silver oxide, the US side uses iodine. The water networks on the Russian and US sides are not connected and the water is supposed to be processed separately. But because not all subsystems have been installed (on the US side, the toilet; on the Russian side, the urine processing unit), and in order to be able to use the US water on the Russian side for oxygen generation, an exchange of water was necessary in 2004. However, mixing iodized water and silver oxide water led to the precipitation of silver iodide, which blocked the pipeline. So a water treatment system had to be developed afterwards to assure safe water transport.

Quality Control of Water

In addition to the sensor system at the end of the water processing chain, the **water quality** is monitored by **measuring conductivity** and pH values and determining iodine or silver oxide concentrations.

5.2.1.8 Fire Detection and Suppression

In order to rapidly detect a fire the modules of a space station are subdivided into individual sections. They are usually in the form of systems or payload racks which are ventilated and separately supplied with smoke sensors. There are also usually additional smoke sensors in the modules themselves and in the ventilation ducts. Besides the removal of heat from the racks, this assures rapid detection of a fire. A fire alarm can also be triggered manually. Fire alarms are indicated by an acoustic warning and a visual indication of fire risk on an LED display. The alarm plan requires the astronauts to immediately put on breathing masks and locate the source of the fire indicated on the display panel. The following measures are carried out either automatically or manually:

- Power is cut off from the affected rack to isolate the fire and turn off its ventilation system.
- **Air exchange** between the affected module and its neighbors is stopped.
- Fire extinguishing measures are initiated.
- The oxygen and nitrogen supplies are turned off.
- The CO_2 removal unit is turned off.
- The analysis units of the **air quality control system** are switched to standby status.
- An evacuation of the module is initiated (venting of the cabin air into space); this can normally only be done by the crew or the mission control center.

In the ISS each rack is a separate **flammable segment**: that is, the racks are separated from each other by bulkheads and fabric. Each rack has one or more small openings in which fire extinguisher nozzles can be inserted so that a rack fire can be quenched locally. The **fire extinguishing chemical** on the ISS is carbon dioxide, stored in conventional fire extinguishers.

Also after a fire it should be possible to make a module habitable again. The CO_2 released from the fire extinguisher is extracted from the cabin air by the CO_2 removal unit. If the module was evacuated to extinguish the fire or remove the gaseous pollutants which were generated, there is the possibility to pressurize it and make it habitable again.

5.2.1.9 Extravehicular Activities

Highlights of extravehicular activities are certainly the repair and refueling of satellites in orbit, the retrieval of satellites for later repair on the ground, and of course the assembly of the ISS in space.

The Airlock

For activities outside a space station or orbiter, or on a planet, space suits are essential. At an internal pressure of 1 bar the suit would be extremely bulky in space and movement not possible, so the internal pressure is only about 300 mbar. To adapt a human body and circulatory system to this lowered air pressure, before exiting into space the astronauts have to spend a certain amount of time in an **airlock** in which the cabin air pressure is reduced to a vacuum or to the atmospheric pressure existing on the planet. The duration is such that the body can adjust to the new environmental situation. The airlock usually has two compartments, one for equipment and the actual airlock. The space suits are stored in the equipment compartment, where some physical exercises can be done, followed by donning or removing the space suit.

In the equipment compartment the astronauts put on **oxygen masks** and begin to breathe pure oxygen. In order to activate their circulatory system, they spend about 10 minutes on a stationary exercise bicycle. Over a period of about 50 minutes the cabin pressure is reduced from 100 to about 30 kPa. During this time nitrogen begins to escape from the blood and leaves the body. At this point the astronauts remove their breathing masks and put on their space suits. Hoses and cables supply the suit and are used to reduce the internal pressure to about 30 mbar and supply it with pure oxygen. The lock is then evacuated to match the outside pressure and the extravehicular adventure can begin.

The Space Suit

The space suit, also called an **extravehicular mobility unit** (EMU), supplies the astronaut with everything required for life support and makes it possible for the astronaut to work in space (Figure 5.2.14). The EMU consists of two subsystems: a life support system and the space suit itself.

The life support system assures:

- The supply of oxygen for breathing and pressure regulation of the space suit.
- Thermal control in the suit through controlled cooling of the circulating water in the space suit envelope.
- Dehumidifying, and removing odors and gaseous pollutants, including CO_2 removal from the circulating oxygen.
- A communication link to and from the astronaut.

Fresh air, in this case pure oxygen, flows into the helmet behind the astronaut's head, passes first over the face and then down into the suit. Exhaled oxygen, carbon dioxide and water vapor are carried along by the flow, which is drawn off at the astronaut's elbows and feet and enters the mobile life support system for processing. This air first passes through an **activated carbon cartridge** to remove gaseous pollutants and then through a **lithium hydroxide cartridge** which removes carbon dioxide. A downstream sublimator cools the air, removing moisture and separating it from the air stream by means of a slurper and a rotating **water separator** before routing it to a water tank. The oxygen flow is then measured and whatever has been consumed is replaced from the oxygen tank. Oxygen circulation is handled by a ventilator (volume flow about 10 m^3/h).

5.2.1.10 Thermal Control

Heat generated by life support aggregates such as the condensing heat exchanger, the electrolyzer, the CO_2 removal system, and especially that generated by the usually water-cooled electronic equipment, has to be eliminated by radiating it to space. Figure 5.2.15 shows a thermal system using the ISS as an example.

A distinction is made between two heat transfer cycles. There is a **low-temperature (LT) bus** with water as the transfer medium, working at flow temperatures between 1 and 6 °C and primarily feeding the condensing heat exchanger by which the air in the module is cooled and dehumidified. The second cycle is the **moderate-temperature (MT) bus**, likewise using water as a heat transfer medium, with a flow temperature of about 18 °C and a maximum return flow temperature of 49 °C. At such temperatures the pipes do not have to be protected against condensate forming on the pipe surfaces. Electronic equipment and other aggregates which are not harmed by these temperatures are cooled with this water. On the ISS a moderate-temperature heat exchanger is simply connected in series downstream of a low-temperature heat exchanger; that is, the moderate-temperature heat exchanger receives as flow the warmed return flow from the low-temperature heat exchanger.

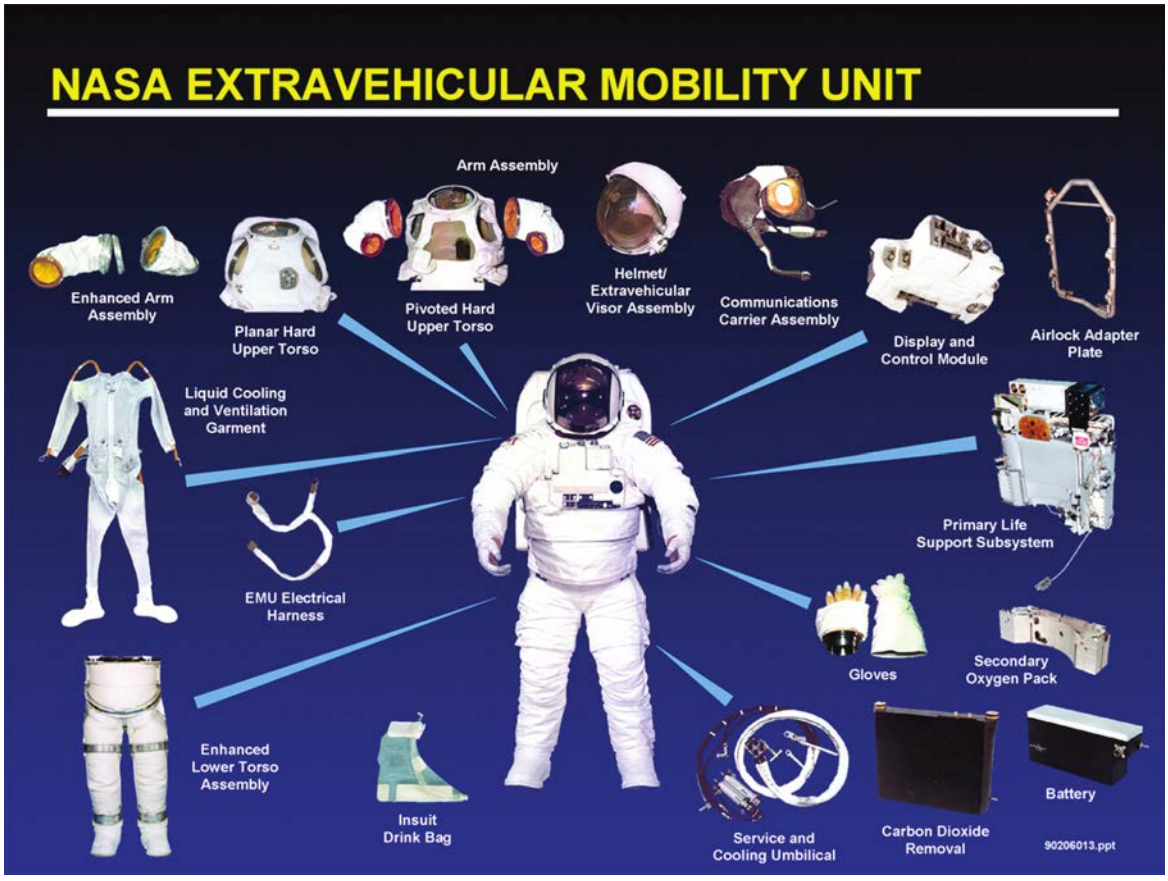


Figure 5.2.14: *Space suit and components (Source: Hamilton Sundstrand Space Systems).*

On the US side of the ISS, these are water–ammonia heat exchangers. The **ammonia** serves as a heat transfer medium for the outer cooling cycle, which ultimately emits all the heat into space via radiators. The ammonia flow temperature is regulated by a three-way valve. In the USOS the direct current transformers are on the outside and the waste heat is directly conveyed to the ammonia. On the Russian side, a brine, **triol**, is used as the internal heat transfer medium, and the refrigerant **Freon** as the heat transfer medium for the outer cycle.

5.2.1.11 Crew Habitation Systems

Crew habitation systems include everything that is directly related to the work and welfare of the astronauts, such as:

- **Personal hygiene**
- Kitchen and **food storage** equipment and functionality
- Collection, recycling and **disposal** of human waste
- **Medical care**
- Personal possessions of the astronauts
- **Mobility aids** and fasteners, attachment points and strips
- Sleeping facilities and restraints
- **Clothing**
- **Radiation monitoring** devices
- Technical guidelines for assuring the maintenance and reparability of facilities and equipment (**handles**, access, maximum **surface temperatures**, protection from electric shocks, avoidance of sharp edges to prevent injury, etc.)

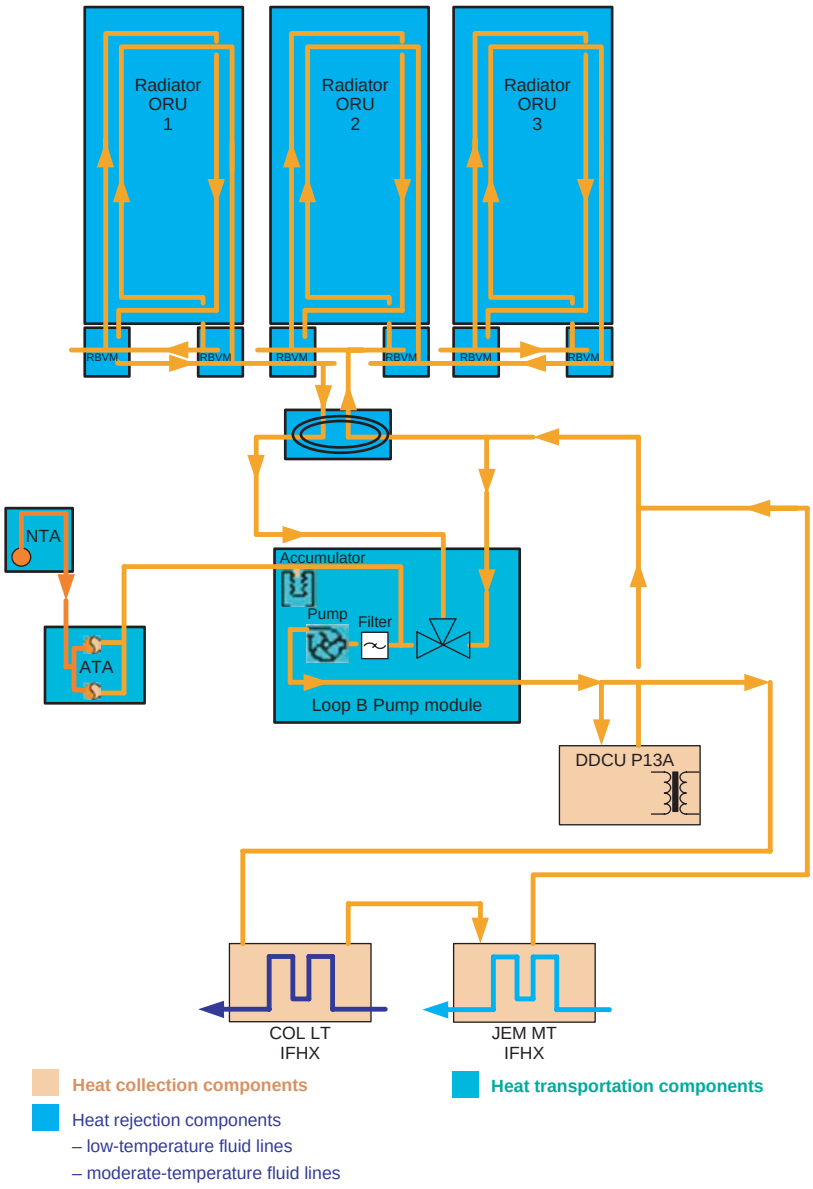


Figure 5.2.15: *Schematic of the cooling water cycle on the ISS.*

- **Noise and odor control**
- Inventory system
- Opportunities to enjoy the outside view
- Facilities and compartments for a private sphere
- Lighting
- Leisure activities
- **Sports equipment.**

5.2.1.12 Food

The first astronauts had to squeeze their meals from tubes; now, after more than 40 years of human space flight, the diet has become much more varied. Eating in a microgravity environment is, however, not a trivial affair. Sugar, salt and pepper cannot be in granular form because they would be too difficult

to put onto food and could be carried away by the air and clog filters. Food that crumbles cannot be offered because the crumbs can get into the eyes and also contaminate the module filters. Since there is no refrigerator for cooling drinking water or food on the ISS, it has to keep well also at temperatures between 20 and 28 °C. The food has to be shaped to conserve as much space as possible to achieve high packing density in the storage lockers.

Astronauts consume four meals a day. In consultation with each individual astronaut, nutrition experts have selected meals that assure that the body receives the proper **balance of calories, vitamins and minerals**. Caloric consumption varies by person and sex between 1900 and 3200 calories. The food consists of fruit, nuts, chicken, beef, seafood and sweets, among many other things, and has usually been conserved by heating, dehydration or radiation. Table 5.2.3 gives an example of a daily menu for the ISS astronauts including a small snack in the evening.

5

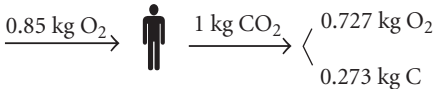
5.2.2 Metabolic Balances

5.2.2.1 An Astronaut's Oxygen Consumption and Carbon Dioxide Production

The amount of oxygen required by a person depends on their size and level of activity. When designing a life support system it is the daily averages which are of interest. Measurements have confirmed the usefulness of the values employed in the design process.

Per person and day an astronaut requires 0.85 kg of oxygen. In order to generate this amount of oxygen by electrolytically splitting water, exactly 0.956 kg of water, about 1 liter per day and person, are required.

A closer look at the human oxygen balance



shows that not all the oxygen is exhaled in the form of CO_2, but that about 15% is metabolized in the body, some of it in water which is later released in the form of perspiration or urine. The so-called **respiration index**

$$R = \frac{O_{2\text{exhaled}}}{O_{2\text{inhaled}}}$$

primarily depends on the fat content of the food and, with the present astronaut nutrition, $R = 0.85 = 0.727\ \text{kg}/0.85\ \text{kg}$.

5.2.2.2 Open Life Support Systems

For brief missions lasting only a few days it is often reasonable to employ **nonregenerative systems** because they are easier to design and represent less of the launch mass. In the simplest case the oxygen is carried in pressurized tanks, cryogenically, or in the form of **oxygen cartridges** and continuously fed into the cabin.

The CO_2 is removed using lithium hydroxide cartridges, and the moisture released to the air by the astronauts is collected with adsorbers like silica gels or a condensing heat exchanger. Typical of open systems is that emitted substances like CO_2, waste water and urine are not recycled but disposed of. Urine is collected and burns up together with an unmanned supply vehicle when it reenters Earth's atmosphere.

5.2.2.3 Closed Life Support Systems

A closed life support system means that waste products on a space station are reprocessed and reenter the cycle. There are different stages in **closing the loops**.

Processing Condensate

Figure 5.2.16 shows that an astronaut drinks about 2.1 l of water per day. Another 0.5 l are contained in the astronaut's food. For personal hygiene an addition 0.2 l are needed. Table 5.2.4 presents a summary.

From Figure 5.2.16 one can also see that 1.5 l of the total of 4.05 l of water consumed can be retrieved via the condensate cycle, for a **regeneration rate** of 37%.

Urine Processing

Each astronaut produces a daily average of 1.3 l of urine. For **flushing toilets** 0.3 l is required. The processing rate is 90%, so $(1.3 + 0.3) \cdot 0.9 = 1.44$ l of water

Table 5.2.3: *The menu of the ISS crew on March 22, 2002.*

Meal 1
Chicken w/eggs (T)
Rossiyskiy cheese (T)
Grape-plum juice (R)
Strawberry tea w/sugar (R)
Meal 2
Pork goulash (T)
Pickled cucumber/meat
Soup (R)
Spiced pike perch (T)
Mashed potatoes/onions (R)
Table bread (IM)
Grape-plum juice (R)
Meal 3
Chicken fajitas (T)
Corn (R)
Tomatoes & eggplant (T)
Candy coated peanuts (NF)
Lemon-lime drink (B)
Meal 4
Cinnamon roll (NF)
Crackers (NF) x2
Lemonade (B) x2

(B) = Beverage
(T) = Thermostabilized
(R) = Rehydratable
(I) = Irradiated
(IM) = Intermediate moisture
(NF) = Natural form

which can be recycled, raising that total regeneration rate to 73%.

Oxygen Cycle

Via a Sabatier reactor, the oxygen cycle can be closed by converting into water and methane (CH_4) two material flows which have not yet been exploited, namely exhaled carbon dioxide and the hydrogen generated by the electrolysis unit:

$$CO_2 + 4\,H_2 \rightarrow CH_4 + 2\,H_2O$$

In order to convert all the hydrogen, stoichiometric considerations require using 58% of the available CO_2. The other 42% of the CO_2 cannot be converted and is discarded to space, as on the ISS. If no additional hydrogen is available, such as from the fuel supply, no more oxygen can be gained from the Sabatier reactor. The water produced amounts to about 0.45 l per astronaut and day, which increases the regeneration rate to 84%.

Hydrogen Cycle

All the CO_2 could be utilized in the Sabatier reactor if the hydrogen bound in the methane could be split off by **methane pyrolysis** and feed into the Sabatier reactor. This would produce an additional 0.32 kg of water per astronaut and day, increasing the regeneration rate to 92%.

5.2.3 The ISS Life Support System

5.2.3.1 ECLS Design Philosophy

Since the life support systems are absolutely critical for a human space station, their design and production have to be of the highest quality. Besides functioning perfectly, there are additional quality requirements like high durability, ease of maintenance and repair, safety, good failure identification, and good man–machine interfaces. Systems must be developed for a service life extending over the entire operational phase of the mission (10 years for the ISS, for example) and which includes the construction phase (five years for the ISS). In order to carry out maintenance and repair work safely and with simple means, maximum surface temperatures cannot be exceeded and requirements have to be met for surface roughness, labeling, the nature of knobs and handles, the design and installation of warning signs, etc.

The associated safety analyses distinguish among failures which are life threatening for the astronauts (**catastrophic**), those which can lead to minor injuries or damage to nearby systems (**critical**) and those with negligible consequences. For example, the failure of a load-bearing structure, the explosion of a pressurized container, and possible contact with high-voltage or unhealthy substances are always ranked as catastrophic. Motivated by safety considerations, it is accordingly specified in the technical design of systems that possible "critical failures" be one-failure tolerant and possible "catastrophic failures" be two-failure tolerant, in order to inhibit their occurrence.

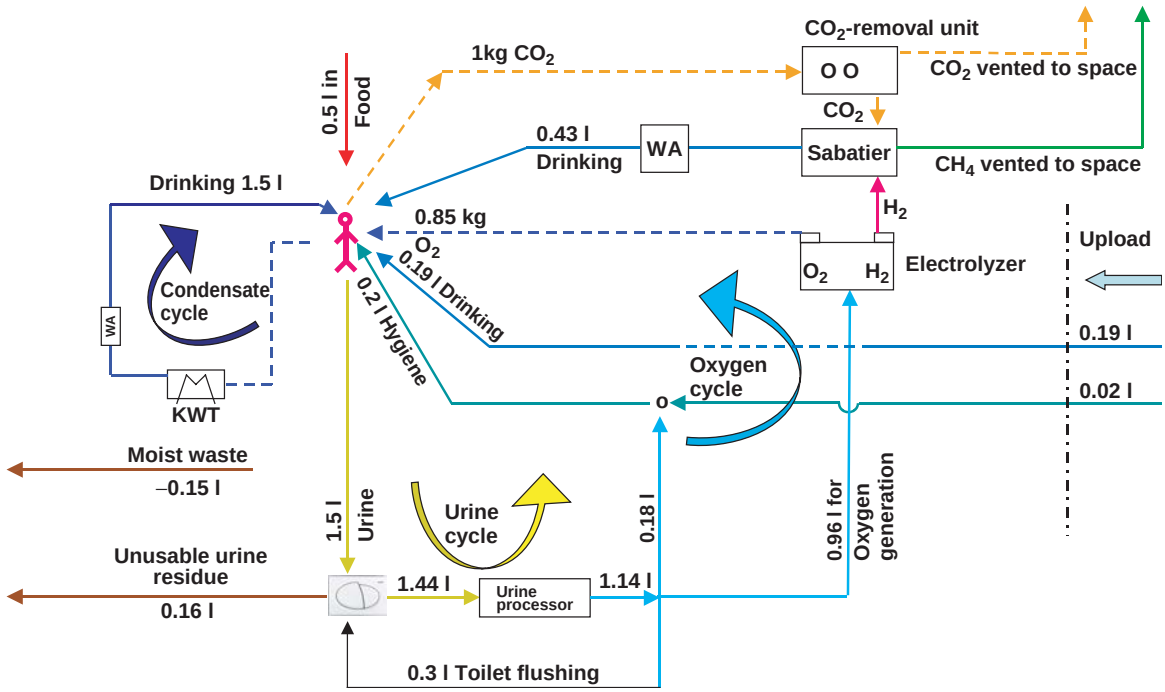


Figure 5.2.16: *Closing the water and oxygen cycle on the ISS.*

5

Table 5.2.4: *Daily water requirements and supply for one astronaut.*

Water requirements (liters/person day)		Water sources (liter/person day)	
Drinking	2.1 l	Condensate	1.5 l
H_2O in food	0.5 l	Urine processing	1.44 l
Hygiene	0.2 l	Sabatier reactor	0.45 l
Toilet flushing	0.3 l	H_2O in food	0.5 l
Oxygen generation	0.95 l	Moisture loss	0.16 l
Total	**4.05** l		**4.05** l

The requirements regarding **fire prevention** are:

1. Manual activation of a fire alarm within 60 seconds of detection.
2. Isolation of a fire may not lead to the loss of a function which could result in a "catastrophic" ranking.
3. Fire extinguishers must be available next to all potential fire sources.
4. The firefighting substance should not damage materials, may not exceed the 1 hour threshold value for gas pollutants, may not be corrosive, and it must be possible to remove from the air any of its by-products with the existing pollutant removal system.
5. Face masks and fire extinguishers must be located at fixed distances throughout the station.

The particular **design philosophy** used to fulfill these requirements may vary considerably. During construction of the ISS it was the US approach to make systems as **redundant** as possible (e.g., to have two identical subsystems, one of which is in operation and the other only switched on if the first fails, or to operate two identical subsystems in parallel, but at reduced power). So in the ISS there are, for example, two CO_2 scrubbers, one located in the US laboratory module and one in Node 3. The Russian approach is to have a **replacement** system available based on a different technological design to be used if a major component fails. For example, if the CO_2 scrubber is

defective, lithium hydroxide cartridges are on hand for CO_2 absorption; if the electrolyzer fails, oxygen cartridges are ignited to produce oxygen until a replacement unit can be provided. This design philosophy is viable because the ISS is regularly resupplied from Progress and Soyuz rockets. For lunar or Mars missions another solution would be necessary.

In connection with a failure-tolerant design, the concept of **design to minimum risk** is used. Liquids that can seriously endanger health and thus represent a "catastrophic" threat for the astronauts have to be locked away with a two-failure-tolerant approach. One could also say that possible contact with this liquid by an astronaut must be hindered by two-failure tolerance. This might be realized, for example, by storing it in two nested containers. If one container fails, the other could still prevent contact with the liquid. But it is also possible to lock away the liquid in only one container, making it so massive (designing it to minimum risk) that its failure according to the state of the art is not plausible.

Most of the equipment in the life support system is zero-failure tolerant: that is, if the unit fails, then its function is not available (the failure of the electrolyzer, for example, would lead to the end of oxygen production with this equipment). However, the design philosophy would then specify either the operation of a duplicate aggregate, or the availability of another unit that can guarantee the performance of the function. Exceptions confirm the rule here, too. For example, the functions of ventilation between modules, ventilation within a module, heat removal and monitoring the module atmosphere are all one-failure tolerant.

There are other differences in design philosophies regarding individual aggregates. For example, the US electrolyzer is composed of several orbital replacement units (ORUs). These are a number of interconnected individual components regarded as one unit and replaced as such in case of failure. In other words, NASA puts a priority on the reparability of equipment. Although the Russians also use water electrolysis for oxygen generation, the Russian electrolyzer on the space station is not designed to be repaired in orbit; it has to be replaced if it fails.

NASA also pays attention to minimizing consumables and using regenerative technologies as much as possible, examples being condensate and urine processing for obtaining drinking water, and oxygen generation using electrolysis.

5.2.3.2 ECLS System

As this handbook went to press, the ISS was still in the construction phase. Figure 5.2.17 shows the planned final configuration, whereby modules not yet in orbit (as of June 2008) are indicated by a dashed line. The modules on the right are part of the USOS; those on the left are part of the **Russian On-Orbit Segment** (ROS). Columbus and JEM are purely research laboratories and contain only minimum basic ECLS equipment. The critical functions of air and water processing are contained in Node 3. An overview of the distribution of the various life support systems as the ISS is configured is shown in Figure 5.2.18.

5.2.3.3 Supplying the ISS and Waste Management

Regular supply flights are necessary to supply the ISS with food, water, oxygen and nitrogen, fuel, consumables, replacement parts and science experiments, as well as astronauts. Table 5.2.5 lists the various **transport systems** currently available or in development.

5.2.4 Biological Life Support Systems

Peter Kern

5.2.4.1 Introduction

The life support systems described above are all based on physicochemical processes. The supplies of oxygen and food for the entire mission were brought up from the ground. Such a procedure is very attractive for missions in low Earth orbit which can rely on supply flights.

For upcoming programs which foresee human long-term missions to the Moon and Mars, the present purely technical life support systems are no longer optimal since they cannot be efficiently, if at all, supported with supply flights – particularly in the case of a Mars mission.

Such scenarios require the use of regenerative, biological life support systems. In practice, a combination

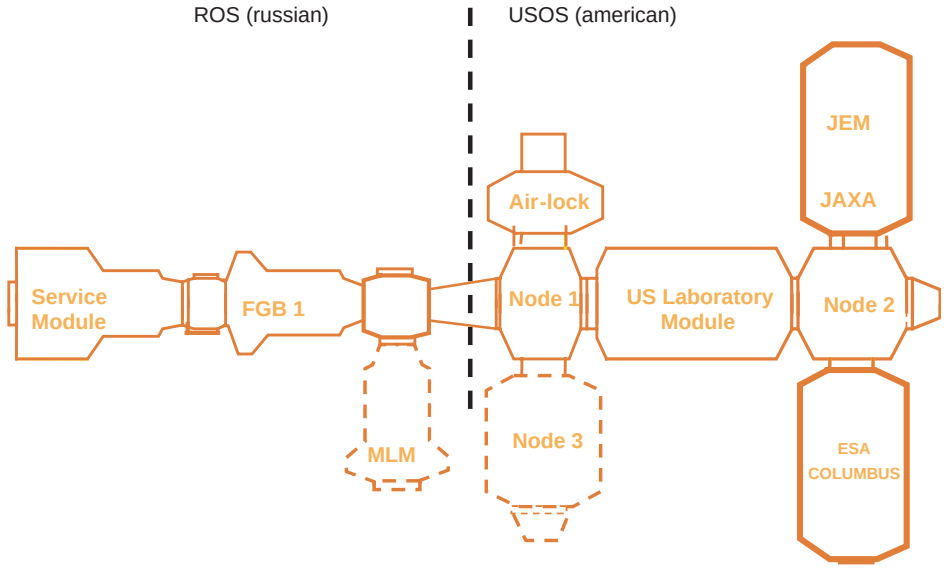


Figure 5.2.17: *Planned final configuration of the ISS.*

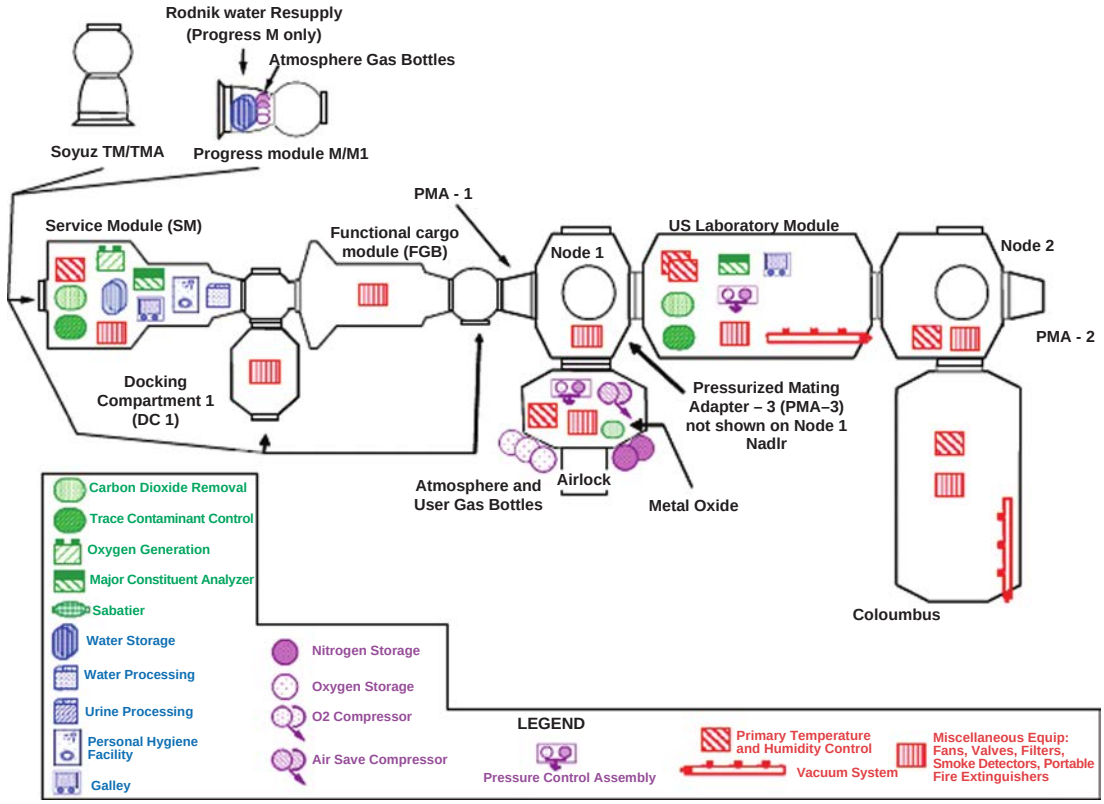


Figure 5.2.18: *Current distribution of the various life support functions on the ISS, as of June 2008.*

Table 5.2.5: *Transport system to and from the ISS.*

Transport system	ISS-Partner	Goods transported	Payload mass [Kg]	Mission interval and duration
Space Shuttle with MPLM or Expresspallet	NASA	Upload and download: • Modules such as US lab (NASA), Columbus (ESA), JEM (JAXA) and Nodes 1, 2, 3 • Four astronauts maximum • Water, O_2, N_2 • Standard supplies • ISPR for experiments and systems • Replacement parts	6979 to 7811	5–6 per year
Progress	RSA	Upload only: • Fuel, food, water, O_2, N_2, replacement parts	2500	5 per year
Soyuz	RSA	Upload and download: • Three astronauts • Approx. 30 kg miscellaneous supplies		2 per year
ATV	ESA	Upload only: • Fuel, water, O_2, N_2, food, replacement parts	6537	Every 15 months
HTV	JAXA	Upload only: • Water, O_2, N_2, food, replacement parts	3090	

of tested physicochemical procedures and specific **biological components** or services will be used. Typical functions which can be filled by biological or biotechnological processes are:

- Removing exhaled CO_2 and generating oxygen via photosynthesis (e.g., by using plants or algae bioreactors).
- Cleansing the air from volatile organic compounds in degassing products or released by metabolites by using **biofilters** or **algae reactors**.
- Food production.
- Processing waste water and urine for direct use or as a feedstock for the biological systems.
- Breaking down, digesting and recycling nonedible parts of food, indigestible waste from humans, or process residue.

There is a considerable amount of literature on the fundamentals of biological life support systems [5.2.2], [5.2.5], [5.2.6], [5.2.7].

5.2.4.2 Terrestrial Developments and Experience at Systems Level

Initial experiments with relatively simple but operationally robust closed ecosystems were carried out in the Soviet Union between 1950 and 1960. They were expanded to include human-populated systems at the Institute of Biophysics in Krasnoyarsk, Siberia, under the leadership of Josef Gitelson in the **BIOS Program**, and continued up to the BIOS-3 facility. Two to three test subjects lived for up to six months in this closed system [5.2.8]. The air was regenerated via photosynthesis in a **chlorella algae reactor** and in up to three plant chambers (**Phytotrons**), which were also used to **produce food**. Protein food such as meat was, however, introduced into the system from outside. The facility was able to recycle waste water and urine. Waste was broken down via catalytic combustion and then added to the biological cycle for reprocessing. Feces were dried but not processed. The cycle was almost, but not completely, closed (approx. 85%).

The only major human-populated facility for a closed biological life support system is **Biosphere-2** in Arizona. It is based on a combination of different climatic zones such as desert, mangrove forest, ocean, rainforest and the astronaut crew habitat, and is complemented and supported by an elaborate operations technology ("Technisphere").

This facility can supply up to eight test subjects with air, food and water in a closed system [5.2.9].

Although there were some technical problems in the initial phase and sometimes an inadequate supply of food, Biosphere-2 is today the longest permanently functioning life support system.

NASA has also intensively investigated this subject since the mid-1980s. The focus is on:

- Food production
- Oxygen production via **photosynthesis**
- Pollution removal using plants and symbiotic microorganisms
- Recycling waste water and urine.

5.2.4.3 Plant Experiments

Since the beginning of space activities many plant experiments have been undertaken to study selected components and processes of biological life support systems.

A good overview of facilities, step-by-step technical improvement and experimental results can be found in Casado [5.2.7]. Relevant equipment on the first Russian space station, Salyut (1971), the MIR missions, and the ISS up to 2006 is covered.

5.2.4.4 Open Questions

Most activities concentrate on the technical and process aspects of life support systems. The biological systems are supported with relatively high technical effort (e.g., MELiSSA, Figure 5.2.19). In most cases their operational stability and tolerance for malfunctions without external control is rather low.

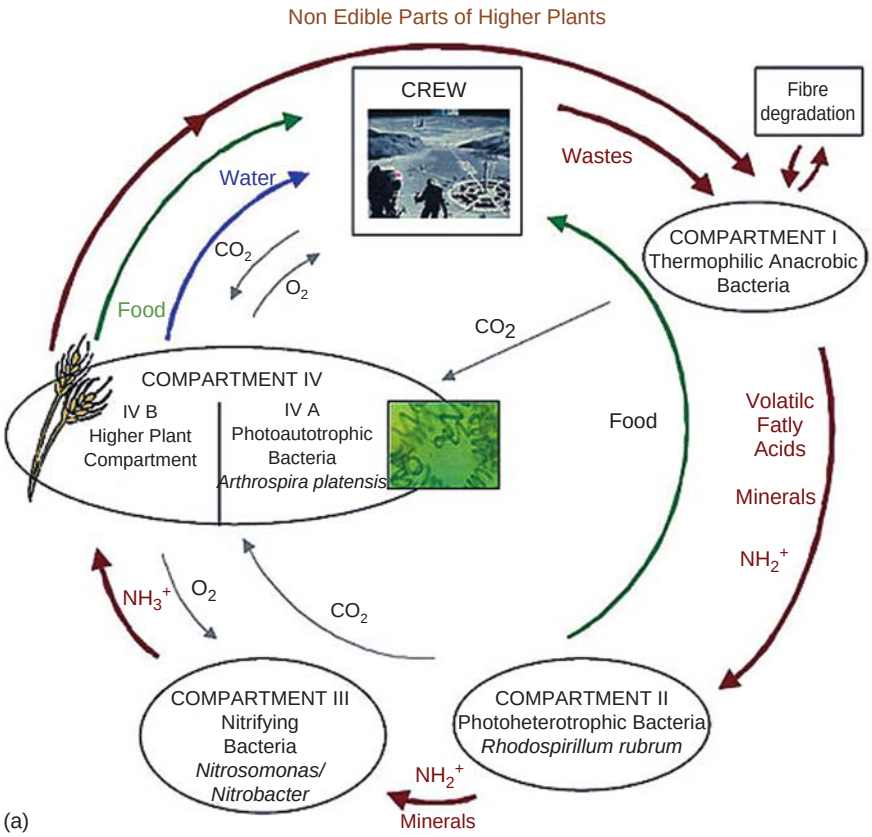
There is less work published about other, operationally related problems with the long-term performance of these systems. Typical examples are as follows.

A: For long-term missions, reliable and operationally robust performance is essential. This is normally achieved by using **redundant** or extremely dependable systems. It is theoretically possible to arrange redundancy for biological components or even entire systems. The following questions arise:

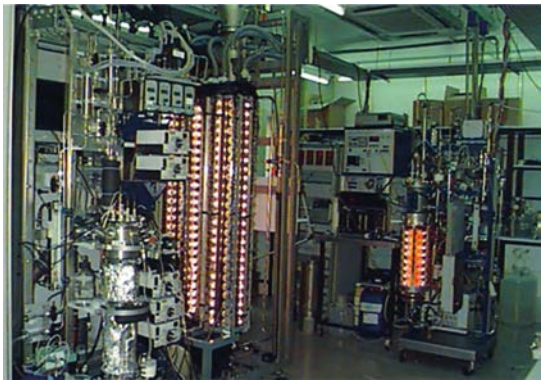
1. While system A is in operation and supplying the crew, what is the **nutrition source** for the redundant system B in the meantime? The crew and the waste products it produces have to be ruled out because, if the primary system A in use were to become contaminated, the redundant system B would also be influenced and no longer available as an optimal, contaminant-free system.
2. But if the redundant system B is running at reduced output, just to keep it operating, then how rapidly can its **performance** be ramped up to efficiently take over the recycling processes? Even a simple biofilter for biological cleansing of used air requires several days to be activated. Aquariums need several weeks of adjustment before their condition is stabilized.
3. How can a contaminated or collapsed biological system in orbit be **cleaned** or sterilized and then reconfigured?
4. How can the **biological starter cultures** needed for the purpose be safely stored in advance so that they can become biologically active enough when required?
5. How can the biological safety of the technical facilities be reliably tested before they are put in operation? It is not sufficient merely to sterilize the equipment. It has to be assured that it does not emit any outgassing residue afterwards, that all materials are compatible with the **sterilization process**, and that the remains of dead biological material such as cell walls are removed by bacteria. These fever-generating sources are tailored to the crew, so they can become very dangerous. Actually, one should strive for the absence of anything causing fevers.

B: During human long-term missions the accumulation of **pollutants and metabolites** in the regenerative environment is a matter of concern. Typical substances are outgassing products from the technical facilities and equipment as well as metabolic degradation products from the biological systems and from nature:

1. For example, the plants themselves produce **ethylene**. As a plant hormone it controls various phases or effects of plant development such as growth, blossoming, fruit maturation and death. Many plants produce ethylene in order to approach or avoid walls (positive or negative thigmotropism). For other plants, ethylene works as a hormone which negatively influences growth [5.2.2]. Other plants require ethylene to initiate the ripening of their fruit.



(a)



(b)

Figure 5.2.19: *MELiSSA functional model showing (a) the basic structure of the regenerative system and (b) the MELiSSA laboratory (Source: ESA).*

2. Another consideration is that the crew will have to increasingly consume **medication**. Typical examples are antibiotics, contraceptives or cytostatic drugs. These pharmaceuticals are not completely broken down in the human body. Some of the metabolites are already a significant problem for water processing and supply on Earth. And on Earth the natural recycling duration is much longer and the options for decomposition much higher.

C: Another important point is controlling **biological contamination** of the air, water, food and surfaces:

1. It is not enough to identify, typify and quantify the contamination as such. It is also not enough to know its population dynamics.
2. From an operational point of view the most important consideration is positively identifying the most suitable methods of **decontamination** and cleaning before use, and verifying their effectiveness. After large-scale application of these methods, their effectiveness must be verified by suitable on-board analysis or diagnostics.
3. For effective decontamination it is not enough only to **kill or deactivate microbiological loads**. Even segments of cell walls from dead biological organisms represent a high health risk for the crew as pyrogens or endotoxins. In the meantime ***in vitro* pyrogen tests** have been developed to determine the objective sensitivity of individuals to such loads.
4. An open question is how to react when it is determined that some of the crew are reacting negatively to the environment and that there is a **health risk**. Such problems cannot be eliminated before the flight by testing, since sensitivity may arise only in the course of the mission as pollutants accumulate.

D: Another open question is the genetic stability of biological systems. Multigeneration experiments with plants suggest that their biological fitness changes over time. This problem is especially critical when designing and using biological components for life support systems. Experiments to clarify this question are being conducted in the equipment on the ISS called the **European modular cultivation system (EMCS)** under precisely controlled environmental conditions.

E: For successful and robust operation of biological components as part of life support systems, the use of several **coupled, synergistic, biological systems** has clear advantages:

1. Their operational reserves and tolerance for malfunctions is higher.
2. External control can be replaced by biological feedback.

F: Another important point is the **psychological acceptability** of the life support systems by the crew. It is reported that the crew on the Russian MIR missions spoke to the plants growing there. It is hard to imagine someone wanting to have a conversation with the manometer of a physicochemical unit.

Bibliography

[5.2.1] Kurmazenko, E.A., Samsonov, N.M., Gavrilov, L.I. *et al.* Offnormal Situations of the Electron-VM Oxygen Generation System Operation Aboard International Space Station. International Conference on Environmental Systems, Rome, Italy, July 11–14, 2005, Paper ICES 2005-01-2803.

[5.2.2] Eckar, P. *Spaceflight Life Support and Biospherics*, Space Technology Library. Torrance, CA: Microcosm Press and Dordrecht: Kluwer Academic, 1996.

[5.2.3] Bagdigian, R., Cloud, D. Status of the International Space Station Regenerative ECLSS Water Recovery and Oxygen Generation Systems. International Conference on Environmental Systems, Rome, Italy, July 11–14, 2005, Paper ICES 2005-01-2779.

[5.2.4] Knorr, W. The FAE Electrolyser Flight Experiment FAVORITE. International Conference on Environmental Systems, Vancouver, British Columbia, Canada, July, 7–10, 2003, Paper ICES 2003-01-2629.

[5.2.5] Anonymous. Regenerative Life Support – Inputs and Outputs. In Space Station Freedom Design Parameters, *NASA SSP 30362*, http://oregonstate.edu/~atwaterj/io.htm, 1990.

[5.2.6] Czupalla, M. Conceptual Design of a Life Support System for Long-Term Human Mars Missions Based on a Bioregenerative Approach. Diplom-Arbeit, Fachhochschule Aachen, 2004.

[5.2.7] Casado, J. Agriculture in Space. *Spaceflight*, **48** (5), 180–189, May 2006.

[5.2.8] Gitelson, I.I., Terskov, I.A., Kovrov, B.G. *et al.*: Long-Term Experiments on Man's Stay in Biological Life-Support System. *Adv. Space Res.*, **9** (8), 65–71, 1989.

[5.2.9] Nelson M., Dumpster, W.F., Silverstone, S. *et al.*: Crop Yields and Light, Energy Efficiency in a Closed Ecological System. Laboratory Biosphere Experiments with Wheat and Sweet Potatoes. *Adv. Space Res.*, **35** (9), 1539–1543, 2005.

5.3 Rendezvous and Docking

Josef Sommer

5.3.1 Introduction

The first rendezvous and docking, that is the first **mating of two Earth-orbiting systems in space**, took place on March 16, 1966, when Neil Armstrong and Dave Scott maneuvered a Gemini capsule toward an unmanned "Agena" vehicle. Since then, numerous docking maneuvers have been performed and with the installation and operation of space stations (Russian, American and today the International Space Station (ISS)) it has become increasingly a standard maneuver for servicing vehicles. Accordingly, **automatic systems** have been developed, implemented and successfully applied, in particular for unmanned servicing vehicles. Though well mastered and "quasi-standard," the maneuver remains risky and is safety critical, as can be seen by numerous small and sometimes large mishaps, like the Progress/MIR accident in June 1997. Maybe this is the reason why **manually controlled docking** is still the preferred method for servicing vehicles like the American Space Transportation System (STS), the Space Shuttle, when visiting the ISS. Moreover, there is a backup system on the space station which can take over control of the approaching vehicle, and in the case of an unmanned vehicle at least prevent collision.

The existing systems are becoming obsolete (in particular because old electronic components are no longer available); they have relatively high power demands and they are heavy and expensive. This is why almost all large space agencies are working on new systems or improvements to increase performance and reduce cost, mass and power consumption. The European Space Agency (ESA) has also developed a new, fully automatic system based on GPS, lasers and cameras and implemented it in an **automated transfer vehicle** (ATV), the European servicing vehicle supporting the operation of the ISS. The first representative vehicle of the planned four ATVs is called *Jules Verne* and it performed a successful automatic docking maneuver with the space station on April 3, 2008.

Mating in space is not only needed for space station servicing but also a key technology for in-orbit satellite repair (Hubble Space Telescope), the assembly of large space systems, as well as for Moon or Mars missions where samples are to be returned to Earth.

The physics of the relative motion of space vehicles in planetary orbits will be briefly described below and concepts for the rendezvous and docking technology will be outlined. In view of the first European successful docking recently performed by *Jules Verne* with the ISS, the focus will be on the ATV concept.

5.3.2 The RVD Mission

The rendezvous typically comprises the steps (a) "phasing," (b) "far-range approach," (c) "final (or short-range) approach" and (d) "docking" (or berthing, if no mating is performed). In the case of a mission to the space station, the servicing is followed by (e) "separation and departure" and finally (f) "deorbiting."

The primary objective of **phasing** is synchronization of the chaser orbit with the target orbit (anomaly, eccentricity, semi-major axis). This is achieved by executing a number of orbit transfer maneuvers, the execution errors of which are corrected by a number of small "trim" maneuvers. Due to the rather long time period between such maneuvers (mainly velocity increments), they may be calculated on the ground with the information then being uploaded by telecommand.

When the desired conditions in terms of position and velocity are reached, the **far-range approach** starts.

The transition point is not fixed but depends on the accuracy and performance of the on-board rendezvous and docking (RVD) system (Figure 5.3.1). Within this phase absolute guidance and navigation is usually replaced by relative guidance and navigation.

At the end of the far-range approach the chasing vehicle is already in the **vicinity of the target** so that additional requirements and constraints such as exclusion zones and approach corridors become applicable. Moreover, the needed accuracy no longer permits ground control but requires **on-board control**. For human systems this can be manual control or automatic control.

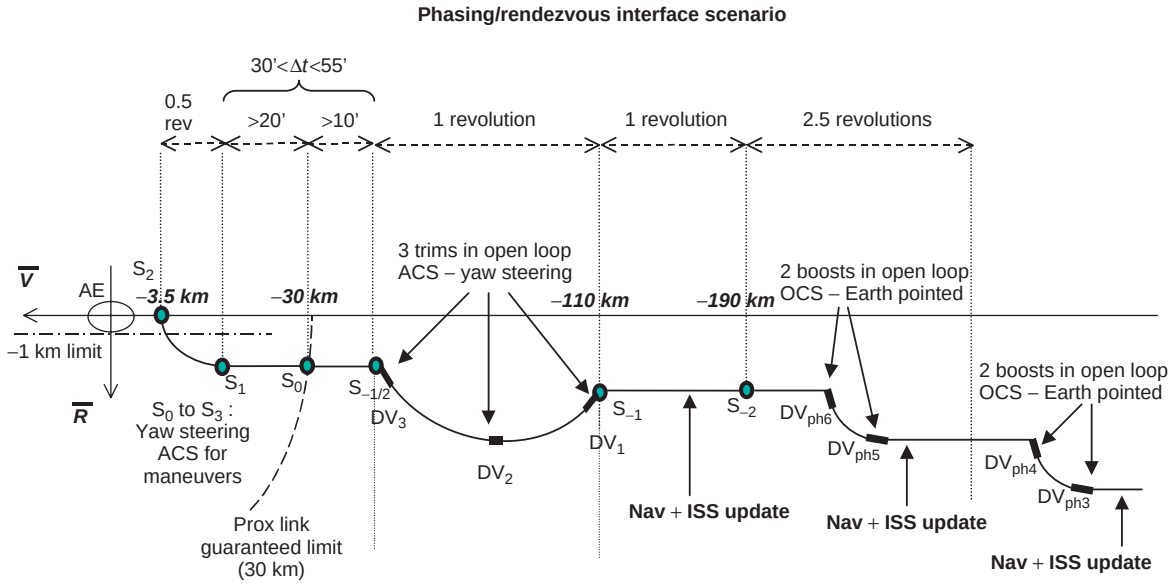


Figure 5.3.1: *The ATV RVD mission. Notation: Δt, time difference; rev, revolution;* ***V****, velocity vector (V-bar);* ***R****, radial vector (R-bar); AE, Approach Ellipse; S, significant point; DV, "delta-v" impulse points; Nav, relative navigation input; ISS, International Space Station position input; ACS, Attitude Control System; OCS, Orbit Control System; Prox link, proximity link for ISS–ATV communication (Source: ESA).*

5

The **docking maneuver** itself depends very much on the docking mechanism. The ISS is equipped with two types of mechanisms, the Russian "probe and drogue" and the American APDS (Androgenous Peripheral Docking System). For unmanned missions such as for a Mars sample return mission, there is no need for pressurization and fluid transfer through the docking port, thus resulting in a much simpler, lighter and cheaper mechanism.

In a rough approximation, **separation and departure** may be regarded as an inverse approach. Departure from the space station must be within a well-defined corridor in a well-defined time frame. Only after sufficiently large distancing can the thrust- and fuel-demanding deorbitation maneuver be executed.

Before presenting some insight into the physics of relative motion, commonalities and differences in the rendezvous approach of today's servicing vehicles for the ISS will be briefly reviewed.

5.3.2.1 Space Shuttle (USA)

For the US Space Transportation System **phasing** is controlled from the ground. It is completed after arriving in the target orbit at a location about 74 km behind the target. Here a thorough checkout of all systems is performed. Thereafter clearance is given for a further approach to about 15 km, this time controlled by on-board systems. The subsequent phase depends on the selected docking port at the station, which can be either in the velocity direction (**V-bar**) or in a direction toward the center of the Earth (**R-bar**). In the case of V-bar, a point about 300 m in front of the docking port is acquired. This is the starting point for so-called "V-bar hopping," an approach method which complies with the safety requirements (see also Section 5.3.3). In the case of an R-bar docking, a braking maneuver is performed at the point where the trajectory crosses the prolonged R-bar axis and then so-called "R-bar hopping" is started toward the docking port (see Figure 5.3.2).

In both cases the cargo bay is oriented toward the target docking port. Though the shuttle is able to perform automatic rendezvous and docking, the last 100 m is usually **manually controlled** by the pilot. This is supported by navigation aids such as a COAS (Crew Optical Alignment Sight, a passive optical instrument), CCTV (Closed Circuit Television

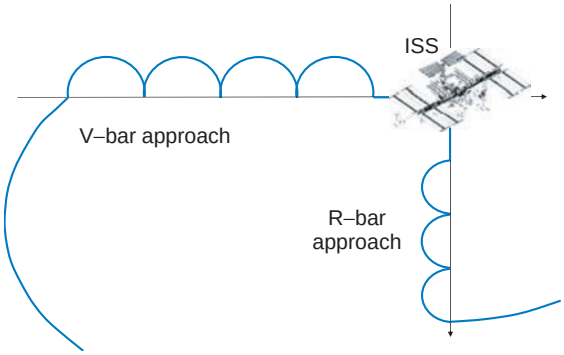


Figure 5.3.2: *Space Shuttle RVD (concept sketch).*

System), consisting of two cameras in the front and aft ends of the cargo bay, and a TCS (Trajectory Control Sensor), which is a laser sensor providing distance and bearing angle measurements.

The shuttle mass is about 100 t. Consequently, strong thrusters are needed for the orbital maneuvers, particularly during phasing and deorbiting. This is provided by the **orbit maneuvering system** (OMS) with its two 26 700 N thrusters. Smaller maneuvers as needed for the proximity operations are performed by the activation of a combination of the 38 RCS (Reaction Control System) thrusters, each providing 3870 N, and the six vernier engines with a thrust level of 110 N each.

Mission planning ensures that the final approach is always performed during daylight.

5.3.2.2 Soyuz/Progress (Russia)

In contrast to the shuttle, the phasing of Soyuz and Progress is not performed in eccentric orbits but essentially in circular orbits. Arrival and departure of these **phasing** orbits are performed by ground-controlled **Hohmann maneuvers**. Appropriate node corrections ensure a coplanar orbit with the target at the end. The far-range approach is performed automatically by the on-board systems and starts with a transfer maneuver toward an aiming point on the target orbit about 1.5 km in front of the target. For navigation the **RF "Kurs" system** is applied, which provides relative distance and velocity as well as the bearing angles.

The **final approach** is performed with the Sun shining from behind (30 to 60°) so that the docking port is well illuminated. If the target is not Earth pointing then the docking port generally is not parallel to the flight direction. In this case a "fly-around" maneuver is introduced to position Soyuz/Progress at a relative distance of 200 to 400 m opposite to the docking port.

The successive final approach is automatically controlled and follows an imposed straight line toward the docking port, ending with mechanical contact at a relative velocity of 0.1 to 0.3 m/s.

Soyuz, with a mass of about 7 t, is equipped with a **main engine** of 3000 N, 14 thrusters of 130 N force for position control (translational motion) and 12 thrusters of 26 N each for attitude control.

5.3.2.3 ATV (Europe)

With the ATV, Europe is providing an important contribution to the servicing of the ISS. The maiden flight with high-precision automatic docking at the Russian docking port (V-bar) was performed on April 3, 2008.

As for the other RVD systems, **phasing** is controlled from the ground. The orbit and maneuver calculations are supported by GPS measurements which the ATV sends to ground. At the end of this phase ATV is in a circular orbit about 6 km below the ISS's altitude and drifting towards the station (see $S_{-1/2}$ in Figure 5.3.1). The far-range approach starts when the **communication link** with the ISS (local link) is established and stable. Via this link ATV receives GPS measurements from the ISS which allow high precision determination of the relative position and velocity needed for the on-board calculation and automatic execution of the maneuvers to reach aiming point S_2, on the target orbit, about 3.5 km behind the station.

This point is left only when permission is given from the space station. After receiving the go-ahead command, the **closing** phase is started and the relative distance is reduced to about 300 m when reaching aiming point S_3. Again, a go-ahead command is needed to proceed with the final approach by means of forced motion in a straight line toward the docking port. When in close vicinity, an optical **RVD camera** is used as the primary navigation sensor.

The orbital maneuvers are performed with the four main engines of 490 N each, while the smaller

maneuvers and attitude control of the 20 t vehicle are performed with the 28 ACS thrusters of 220 N each.

5.3.3 Basics of Relative Motion

A detailed description of orbital mechanics, in particular for the relative motion of two vehicles in circular low Earth orbits, is given in [5.3.1]. In this section, only a short synthesis of the equations of motion and their impact on the selection of approach strategies will be given. The latter are driven by safety requirements, mainly collision avoidance, which needs to be ensured even in cases of loss of vehicle control.

When using a reference system with its origin in the target's center of mass (Figure 5.3.3), the relative motion in a circular orbit can be described by a set of simple **linear differential equations** known as the Euler–Hill (sometimes also called Clohessy and Wiltshire) equations:

$$\begin{aligned} \ddot{x} &= \frac{F_x}{m} + 2\,\omega\,\dot{z} \\ \ddot{y} &= \frac{F_y}{m} - \omega^2\,y \qquad\qquad \omega = \frac{2\,\pi}{T_{\mathrm{Orbit}}} \\ \ddot{z} &= \frac{F_z}{m} - 2\,\omega\,\dot{x} + 3\,\omega^2\,z \end{aligned} \tag{5.3.1}$$

where T_{Orbit} is the time needed by the target for one orbital revolution, m is the mass of the chaser vehicle, and F_k are the components of the sum of the forces acting on the chaser (external perturbations like air drag and forces generated by the thrusters).

Looking at these equations we can see immediately that the motion is decoupled into an in-plane motion (x–z) and an out-of-plane motion (y). An **analytical solution** of these equations is given in [5.3.1]. Assuming **constant acceleration** over the time of interest, the solution is

$$\begin{aligned} x(t) &= \left(\frac{4\,\dot{x}_0}{\omega} - 6\,z_0\right)\sin(\omega\,t) - \frac{2\,\dot{z}_0}{\omega}\cos(\omega\,t) \\ &\quad + (6\,\omega\,z_0 - 3\,\dot{x}_0)\,t + \left(x_0 + \frac{2\,\dot{z}_0}{\omega}\right) \\ &\quad + \gamma_z\,\frac{2}{\omega^2}\left[\omega\,t - \sin(\omega\,t)\right] \\ &\quad + \gamma_x\left\{\frac{4}{\omega^2}\left[1 - \cos(\omega\,t)\right] - \frac{3}{2}\,t^2\right\} \\ z(t) &= \left(\frac{2\,\dot{x}_0}{\omega} - 3\,z_0\right)\cos(\omega\,t) + \frac{\dot{z}_0}{\omega}\sin(\omega\,t) \\ &\quad + \left(4\,z_0 - \frac{2\,\dot{x}_0}{\omega}\right) + \gamma_x\,\frac{2}{\omega^2}\left[\sin(\omega\,t) - \omega\,t\right] \\ &\quad + \frac{\gamma_z}{\omega^2}\left[1 - \cos(\omega\,t)\right] \\ y(t) &= y_0\,\cos(\omega\,t) + \frac{\dot{y}_0}{\omega}\sin(\omega\,t) \\ &\quad + \frac{\gamma_y}{\omega^2}\left[1 - \cos(\omega\,t)\right] \end{aligned} \tag{5.3.2}$$

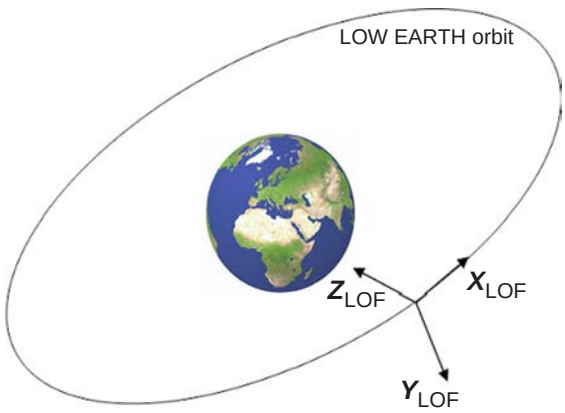


Figure 5.3.3: *Earth-pointing coordinate system (rotating; LOF, Local Orbital Frame).*

For the out-of-plane motion the result is an **undamped oscillation** with orbital pulsation. The amplitude is determined by the initial conditions for out-of-plane position and velocity by

$$y_{\max} = \sqrt{\left(\frac{\dot{y}_0^2}{\omega^2} + y_0^2\right)} \tag{5.3.3}$$

The motion within the orbital plane corresponds to a **cycloid**, the radius of which is given by the difference in the semi-major axis of the target and chaser. The describing point of the cycloid is determined by the **eccentricity of the chaser orbit** (note that the target orbit is assumed to be circular).

According to [5.3.2], when setting external forces to zero ($\gamma_i = 0$) these equations correspond to those for a **walking ellipse** (Figure 5.3.4) with a semi-major axis twice as large as the semi-minor axis and the center coordinates given by

$$x_C = x_0 + \frac{2\,\dot{z}_0}{\omega} \qquad z_C = \frac{2\,\dot{x}_0}{\omega} - 4\,z_0 \tag{5.3.4}$$

The walking ellipse moves in the x-direction with a velocity of

$$v_C = 6\,\omega\,z_0 - 3\,\dot{x}_0 \tag{5.3.5}$$

The **semi-minor axis** is given by

$$b = \sqrt{\left[\frac{\dot{z}_0^2}{\omega^2} + \left(2\frac{\dot{x}_0}{\omega} - 3\,z_0\right)^2\right]} \tag{5.3.6}$$

The **semi-major axis** is then $a = 2b$.

With $z_C = 0$ (target and chaser on the same orbit) the **drift velocity** is also zero ($v_C = 0$). The result is motion along an ellipse at a distance to the target of x_C. For $z_C > 0$ (chaser below the target orbit), on average the chaser is moving faster than the target; for $z_C < 0$ (chaser above the target orbit) the chaser is slower. Figure 5.3.5 shows the typical motion for both cases for an initial condition of $z_0 = 50$ and -50 m, respectively, and a radial velocity of $\dot{z}_0 = 0.3$ and -0.3 m/s, respectively.

If $b = 0$, then we are on a **circular orbit** (no eccentricity) and the drift velocity is given by $v_C = 1.5\,z_0$.

If the target and chaser are also at the same altitude ($z_0 = 0$) then we are at a **hold point** in the target orbit (no relative motion with respect to the target), which

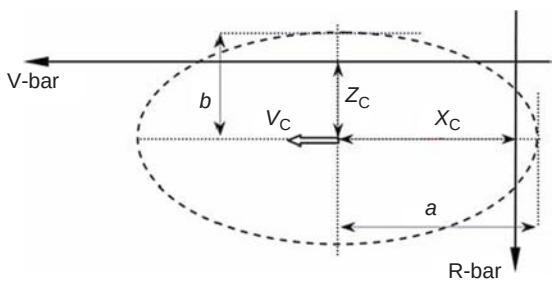


Figure 5.3.4: *Parameters of the walking ellipse.*

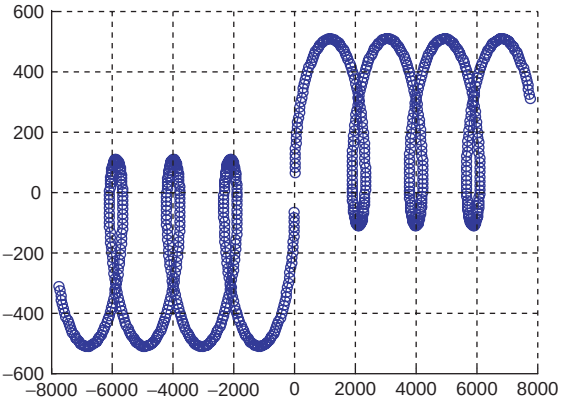


Figure 5.3.5: *In-plane relative motion.*

is the ideal position for awaiting convenient docking conditions and performing a system checkout.

Though the driving element for the **approach strategy** is safety, a suitable approach must also meet visibility constraints (communication, sensor field of view) and comply with the maneuverability limitations of the vehicle. The goal is to find a trajectory which avoids collision even in the case of loss of vehicle control, taking uncertainties in the initial conditions into account. This means that, whenever the approach is aborted for whatever reason, the resulting trajectory should have a negative average drift, that is, the distance to the target should increase. The applied principle is easily understandable when looking at the cycloid of Figure 5.3.6.

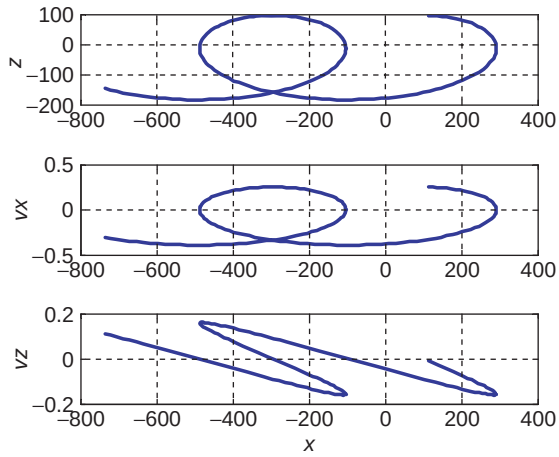


Figure 5.3.6: *Free drift trajectory.*

Let us assume we are at point (−480/0) to the left of the origin of target center (0/0).

When imposing a radial velocity impulse of ΔVz_1 we arrive at the blue trajectory. Though in the short term the chaser is approaching, the motion is reversed at about −100 m and in the long term (average per v_c orbit) the distance is increased. This is because v_c is negative. If the second maneuver ΔVz_2 is not applied or only partially applied, then we are on a collision-free trajectory. If the maneuver is correctly executed we are on a hold point at a closer distance to the target; that is, we performed an approach on a safe trajectory.

A sequence of such maneuvers is called **V-bar hopping**, as above. It should be noted here that the distance between the individual hops can be varied: they may become smaller, the closer we are to the target.

The same principle may be applied for the radial motion. Let us assume we are at point 2 (−300/−160). We arrive at point 1 (−300/+100) along the blue trajectory when imposing a velocity increment in the negative "x" (−0.33 m/s) and positive "z" (0.1 m/s) directions. The resulting motion from bottom to top can be stopped at point 1 by an appropriate maneuver, or continued with a second hop. Accordingly a sequence of such maneuvers is called **R-bar hopping**.

Both strategies can be applied by the Space Shuttle when servicing the ISS (see also Figure 5.3.2).

If the acceleration during a maneuver is high enough, then the changes in velocity can be performed within a short time period and considered as approximately instantaneous. This allows the analytical solution of the Euler–Hill equations to be applied, with $\gamma_I = 0$ for the calculation of the maneuvers (now velocity increments). For this the equations of motion are written in **state space form** as given below:

$$\vec{x}(t_E) = \begin{bmatrix} \phi_{11}(t_E, t_0) & \phi_{12}(t_E, t_0) \\ \phi_{21}(t_E, t_0) & \phi_{22}(t_E, t_0) \end{bmatrix} \vec{x}(t_0) \qquad (5.3.7)$$

$$\vec{x} = [x, z, \dot{x}, \dot{z}]^T ,$$

If the initial position, the desired final position and the transfer duration ($t_E - t_0$) are known, then the necessary **initial velocity** can be calculated via the transition matrix. The required velocity increment then is the difference between the known initial velocity and the computed, required one according to

$$\begin{bmatrix} \Delta V_x \\ \Delta V_z \end{bmatrix} = \phi_{12}^{-1} \left(\begin{bmatrix} x_E \\ z_E \end{bmatrix} - \phi_{11} \begin{bmatrix} x_0 \\ z_0 \end{bmatrix} \right) - \begin{bmatrix} \dot{x}_0 \\ \dot{z}_0 \end{bmatrix} \qquad (5.3.8)$$

The second impulse of this two-burn transfer is determined by the difference between the velocity at the end of the transfer duration and the required velocity in the aiming point (which is zero for a hold point). In the same way the well-known energy-optimized **Hohmann transfer** between two circular orbits can be approximated when half the time of an orbital revolution is taken as the transfer duration. For controlling the out-of-plane motion the same principle can be applied.

The desired velocity increments are realized by appropriate **commands to the thrusters**. Since the force level of the thrusters is (more or less) constant over the duration of the maneuver, the required velocity increment can be transformed approximately to a burn duration by the following equation:

$$t_{Burn} = m \frac{\Delta v}{F_{Thrust}} \qquad (5.3.9)$$

The corresponding fuel consumption is derived from

$$m_b = m_0 \, e^{\frac{-\Delta v}{I_{SP} \, g_0}} \qquad (5.3.10)$$

with m_0 the vehicle mass before the maneuver, m_b the vehicle mass after the maneuver, g_0 the gravitational constant and I_{SP} the specific impulse of the applied thruster.

If only small thrust levels (more precisely, small accelerations) are available, the approximation given above for maneuver calculations results in large errors. In this case the analytical solution with $\gamma_I \neq 0$ has to be applied. The transfer calculation is then split into three segments, the two burn durations and the drift period between them (see also [5.3.3]). A special case is given for constant acceleration over a complete orbital revolution in either the tangential or radial direction.

The application of a constant **tangential acceleration** over one orbit results in an orbit transfer similar to the Hohmann transfer. When starting on a circular orbit, another circular orbit of different altitude is reached at the end of the transfer. The difference in altitude Δz is determined by the acceleration level

$$\gamma_x = \frac{-\omega^2 \Delta z}{4\ \pi} \tag{5.3.11}$$

A constant acceleration in the radial direction is similar to a V-bar hop, when it is applied over one orbital revolution. Here the acceleration level determines the transfer distance Δx:

$$\gamma_z = \frac{\omega^2 \Delta x}{4\ \pi} \tag{5.3.12}$$

In both cases the integral of the accelerations corresponds to the velocity increment which would be needed for an impulsive transfer maneuver as well; that is, there is no difference in maneuver energy.

Impact of Air Drag

For low Earth orbits, the Earth's atmosphere cannot be neglected when considering the motion over a long duration (more than one orbital revolution). The air density (the order of magnitude is $\rho = 10^{-11}$ kg/m^3 at 400 km altitude) not only is a function of altitude but varies according to the solar activity (F10.7) and the geomagnetic index. The resulting **air drag** is given by

$$a_{\text{air}} = \frac{A\ c_{\text{D}}}{m}\ \frac{\rho}{2}\ v^2 \tag{5.3.13}$$

where A is the effective area of the vehicle in the velocity direction, m the vehicle mass and c_{D} the drag coefficient. For the relative motion the difference in air drag between target and chaser needs to be computed. A typical impact over several orbital revolutions is shown in Figure 5.3.7 for an ATV (chaser) and ISS (target) configuration. It can be seen that the initial negative drift (away from the ISS) is reversed after three revolutions to a positive drift (approaching the ISS), thus becoming a potential safety issue.

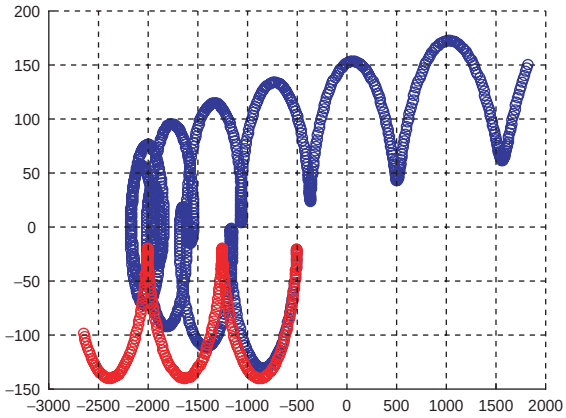


Figure *5.3.7: Free drift trajectory with (blue) and without (red) air drag.*

To a first-order approximation the **air drag** is constant over one orbital revolution. Consequently, it can easily be accounted for in the acceleration γ of the analytical solution of the Euler–Hill equations. It should be noted, however, that only the difference in the air drag effect between target and chaser is used.

5.3.4 The Safety Requirements

The interfaces and requirements for vehicles visiting the space station are defined in [5.3.2] (note that not all requirements apply for existing systems). Figure 5.3.8 shows the safety areas within the orbital plane around the station.

The **approach ellipsoid** (AE) may only be entered if permission is given from the space station ground

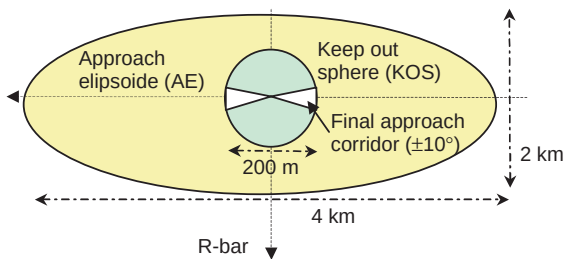


Figure 5.3.8: *ISS safety zones.*

control center. For safety monitoring it is therefore required that the visiting vehicle be on a well-known (and well-predictable) position or trajectory 90 min (about one orbital revolution) prior to arrival at the edge of the AE. The point at which the entrance to the AE is initiated is called the AI (Approach Initiation).

Before the AI is triggered, the vehicle status has to be such that any free drifting trajectories resulting from a potential loss of control of the vehicle (including 3σ uncertainties) stay outside the AE for at least 24 h.

Within the **keep out sphere** (KOS) the vehicle is only allowed to move within a cone of 10° with respect to the desired docking port.

On the request of the ISS, the vehicle must be capable of aborting the approach and withdrawing to a safe hold point. After this, up to 34 h (one day + one ISS crew shift) needs to be bridged before permission for a new approach can be expected. The total duration of the maneuver within the AE may not exceed 10 h, corresponding to one **crew shift**.

When departing from the station a monotonic increasing of the distance is required. The edge of the AE has to be reached after 90 min at the latest and reentrance thereafter is not allowed.

The vehicle must be able to recognize an impending collision and automatically trigger a **collision avoidance maneuver** (CAM). By this maneuver it must be ensured that the vehicle stays out of the AE for at least 24 hours. Should the CAM be executed outside the KOS, then it is not allowed to enter this sphere.

Apart from these operational and mission requirements, there are a number of design requirements to ensure ISS safety, an important one being the required **two-failure tolerance** with respect to collision. This means that any two independent failures within the system or its control unit should not result in a collision risk, or, in other words, after any two failures the system must be able to execute a CAM.

For monitoring and **communicating** with the visiting vehicle the space station provides a bidirectional link, a so-called "proximity link."

Figure 5.3.9 shows the communication zone when approaching the Russian docking port at the V-bar side. This link is used for continuous transfer of data showing the system's health state of the approaching vehicle.

5.3.5 The ATV RVD System

Based on the fundamentals described above, the ATV approach strategy is easily understood (see, e.g., Figure 5.3.10). The AI point corresponds to S_2 in Figure 5.3.11. The nominal coordinates have been selected such that it is about 100 m above the target orbit. Though this increases fuel consumption for station keeping, it improves passive safety because it ensures a negative drift (away from the ISS) if the system control fails. The transition from S_1 to S_2 is called "homing" and corresponds to two impulsive Hohmann transfers with two intermediate correction maneuvers. The latter are needed because of errors in both measurements and navigation as well as in the execution of the maneuvers. The required velocity increments for all maneuvers can be computed according to the formulas given above.

When clearance is given for the rendezvous, the "closing" from S_2 to S_3 is performed by applying a V-bar hop. As for homing, the transfer includes two correction maneuvers and aims at a hold point outside the KOS in front of the approach cone. It is worth mentioning that by selecting a transfer duration of less than half an orbital revolution, a trajectory is established which follows the passive safety concept; that is, in case of loss of control the free flight trajectory will have an average drift away from the station. Like S_2, hold point S_3 represents a time-flexible element in overall mission planning.

For the further approach the ATV must stay within the 10° cone zone extending from the docking port (approach corridor). This essentially prohibits the application of orbital arc maneuvers (hopping). Consequently, the final approach is performed by imposing closed-loop-controlled direct motion (straight line) toward the docking port.

The **approach velocity** follows a profile as given in Figure 5.3.12. It gets smaller as the distance to the docking port closes. The linear decrease with distance is again determined from safety considerations. It ensures that the bulge of a potential free drift trajectory (see also Figure 5.3.6) will not hit the structure

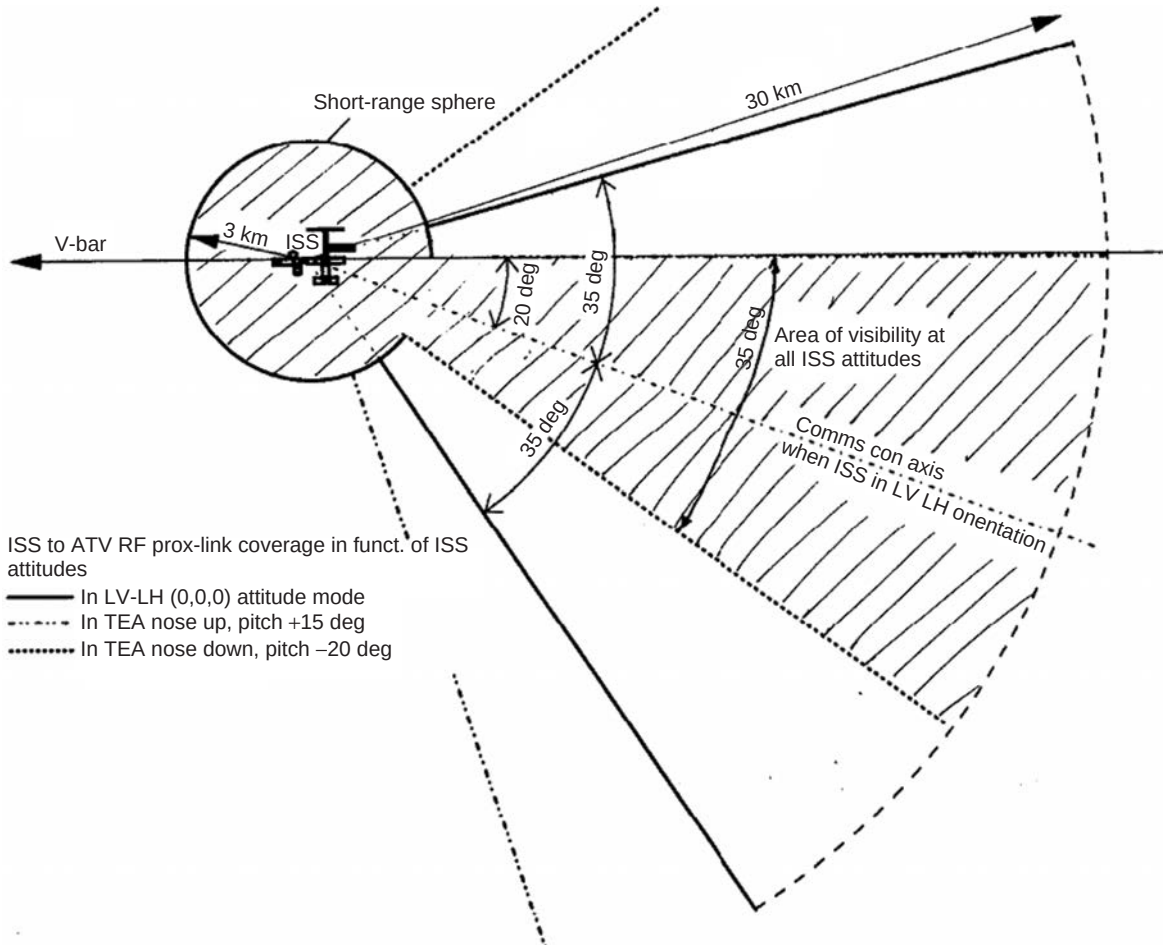


Figure 5.3.9: *ISS proximity link.*



Figure: 5.3.10: *ATV rendezvous with the ISS (Source: ESA).*

envelope of the ISS. At about 40 m relative distance the velocity corresponds to the desired docking velocity (Figure 5.5.13).

Up to point S_4 the trajectory is aligned toward the average (constant) attitude of the docking port. In reality, however, the docking port is moving. This is because the space station also suffers perturbing torques due to air drag, a gravity gradient and Earth's magnetic field which need to be controlled by an appropriate **attitude control** system. Following the resulting motion the last meter of approach needs to be performed in a docking-port-fixed reference system, generally moving with respect to the average frame used before. When synchronized with the ISS attitude motion, the ATV is accelerated toward the docking port up to the desired

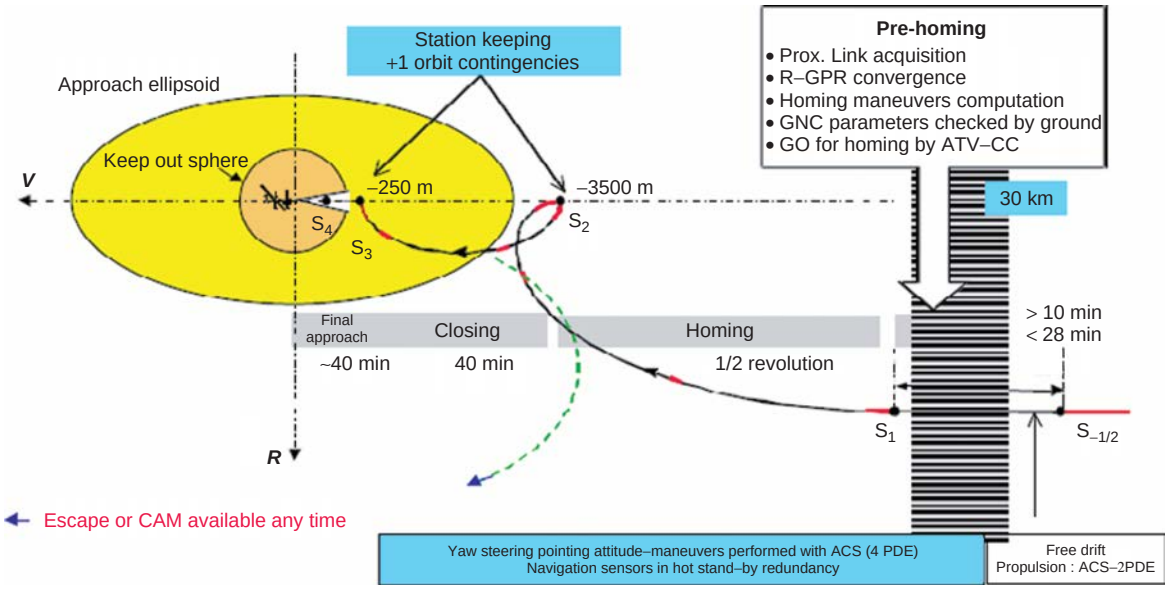


Figure: 5.3.11: *ATV proximity operations (from [5.3.5]).*

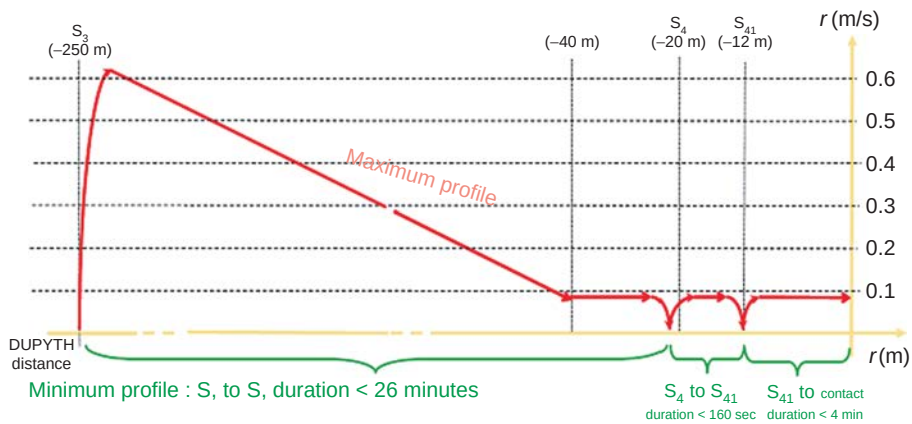


Figure: 5.3.12: *Velocity profile.*

docking velocity of 7 cm/s, which is kept constant until contact is established (Figure 5.3.14). This velocity complies with the needs of the Russian **"probe and drogue" docking mechanism**.

After the first mechanical contact detected by sensors in the probe head, the aft ATV thrusters are fired (post-contact firing) for a short time to avoid rebouncing; the coupled system is damped and the probe retracted to allow mating and connection of interface lines. At the end of the docking procedure the inner part is removed (hatch opened) and the pressurized module of the ATV can be entered from within the station.

The nominal duration of ATV servicing is scheduled to last six months. Besides delivery of materials and consumables (fuel, water, oxygen), the ATV performs boost maneuvers for the space station. After stowage of all disposables, the ATV is prepared for **departure**. The system is booted, the hatch is closed and separation is commanded. An initial separation

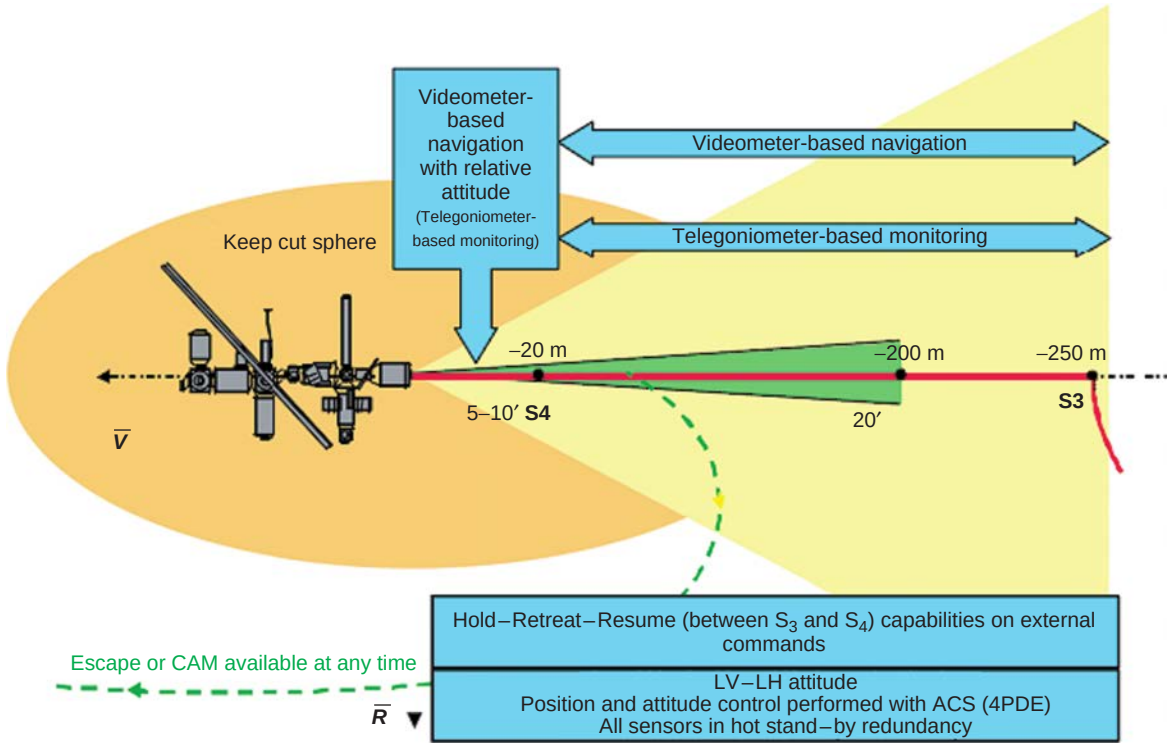


Figure 5.3.13: *ATV final approach (from [5.3.5]).*

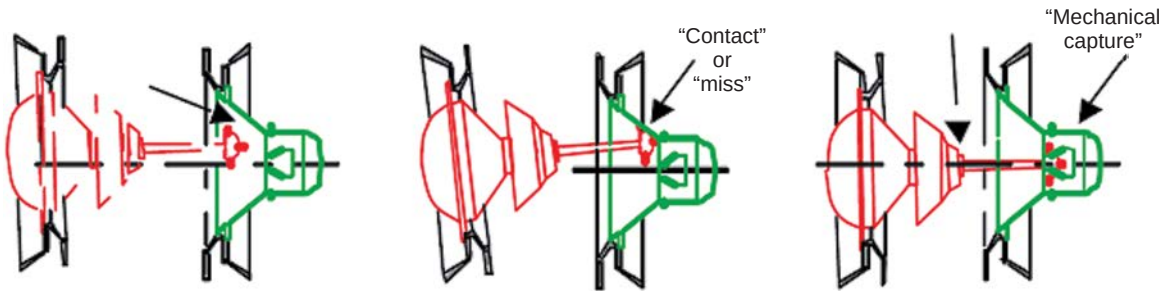


Figure 5.3.14: *Docking sequence.*

impulse is given by the docking port spring mechanism. To leave the AE within the required time a braking maneuver is needed to enter a trajectory similar to the green dashed line of Figure 5.3.13. Subsequently, a large maneuver for transition to the deorbitation orbit is required, which finally ends in a destructive reentry where most of the components burn up in the atmosphere. Only rather compact elements like batteries reach the Earth surface and splash down in the South Pacific.

The requirement for **two-failure** tolerance for the CAM yields rather complex cross-strapped systems

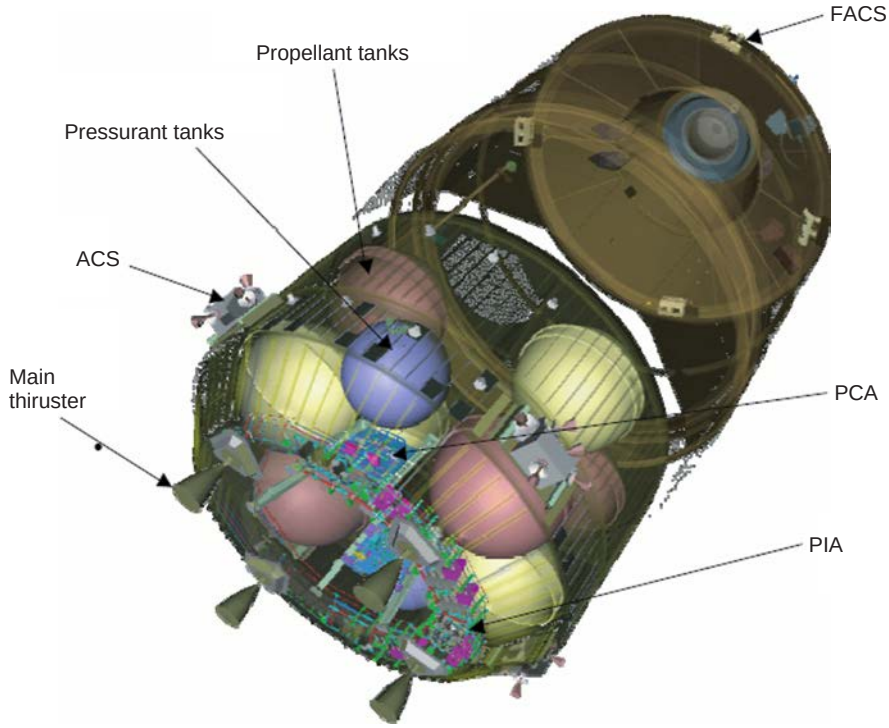


Figure 5.3.15: *ATV propulsion system (Source: ESA).*

with sophisticated monitoring and safety elements. The operational system performing rendezvous is **one-failure tolerant**: that is, after any failure the system is still able to dock. This includes the capability to execute an escape maneuver (similar to CAM). This approach allows implementation of the safety system also with only one-failure tolerance because the first failure is handled within the operational system.

The avionics within the ATV accordingly consist of three independent computers running in parallel and monitoring themselves by appropriate voting techniques. In the case of an error, the failed computer is deactivated. Since it is not possible with two units to identify a potentially failed one (no more voting majority), safety needs to be ensured by another independent system. This is done by the **monitoring and safety unit** (MSU). It consists of two additional computers running in hot redundancy with a dedicated power supply and with dedicated access to the propulsion system. In addition it has an independent link to the TM/TC and receives attitude data from an independent set of sensors.

5.3.5.1 The Propulsion System

The ATV propulsion system (Figure 5.3.15) employs two types of bipropellant thrusters (MON, MMH), the attitude control thrusters (ACT) with 220 N each, and the orbit control thruster (OCT) with a thrust level of 490 N.

The **four main engines of 490 N** are accommodated at the aft part of the vehicle and are co-aligned with the longitudinal axis. Orbit transfer maneuvers of up to 75 m/s require burn durations of more than 30 min resulting in a fuel consumption of about 500 kg ($I_{SP} = 311$).

The eight **tanks** (2×4) allow the stowage of up to 6960 kg of fuel, more than half (4 t) being foreseen to boost the space station. These maneuvers are needed to compensate for the air drag effect, which results in a small but continuous deceleration, invoking a small but continuous decrease in altitude. The two helium tanks with an initial pressure of 310 bar act as a pressure reservoir which allows the operational tank pressure to be kept at a constant value of 18 bar, which is needed to ensure a constant thrust level.

There are 28 **attitude control thrusters** distributed in four clusters (each equipped with five thrusters) on the resource module aft (ACS) and four clusters (each equipped with two thrusters) on the pressurized module at the front (FACS). They are controlled by four propulsion drive electronic (PDE) units such that after any failure, fully six degrees of freedom control capability is guaranteed, and after any two failures a CAM braking maneuver is possible. For the latter at least four thrusters are required along the *x*-axis. This is why each PDE provides dedicated drive electronics for two thrusters out of the eight which can be used for a CAM.

To achieve the required accuracy in control, in particular during docking, the thruster must be able to execute very small impulses. The minimal duration for the 220 N attitude control thrusters is 25 ms. In this operational mode the specific impulse (nominally 286 s) decreases significantly (about 180 s), thus increasing the relative fuel consumption.

5.3.5.2 Guidance and Control

The guidance, navigation and control (GNC) system aboard the ATV determines the dynamic state (position, velocity, attitude and attitude rate) of the ATV, computes the trajectory along which the vehicle is to move, generates the set points for the attitude to be established, and ensures that deviations due to disturbances and perturbations are counteracted. The basic principles of guidance are described in Section 5.3.3. In close proximity the trajectories are composed of segments following simple laws like constant acceleration, constant velocity or linear variation of the velocity as a function of distance (see also Figure 5.3.12).

Four solar arrays are available for **power generation**. Since they can be rotated only around one axis, the ATV needs to establish a certain attitude to allow the solar array area vector to point directly toward the Sun (best efficiency for power generation). The attitude allowing this while keeping one axis pointing to Earth is called "yaw steering." Figure 5.3.16 shows the principle. The ATV rotates around the *z*-axis (pointing towards Earth) such that a second rotation can be performed by the solar arrays, allowing Sun pointing of the latter.

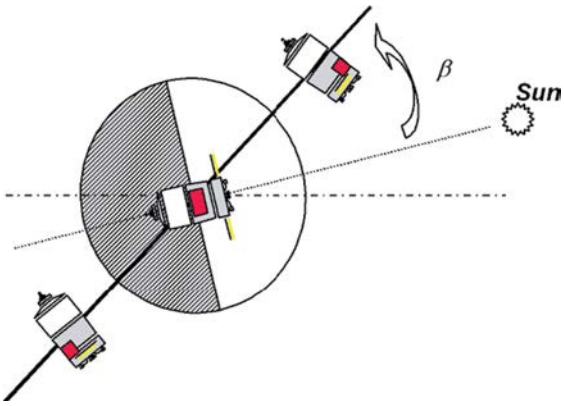


Figure 5.3.16: *Yaw steering.*

For small solar elevation angles β (this angle between the orbital plane and the Sun is a function of season), high angular rates would be requested around *z*. Therefore this concept is applied only for $\beta > 10°$.

The state estimation within the navigation function applies Kalman filter techniques for the optimal determination of position, velocity, attitude and attitude rate, both absolute and relative. With the exception of the last meter, translational and rotational motions are decoupled.

The attitude estimation is based on four gyros, each of which provides a two degrees of freedom measurement. Since gyros are affected by a drift effect, the estimation is regularly updated by means of star sensor measurements. To avoid singularities (also known as gimbal lock) the attitude kinematics are formulated in terms of quaternions (see [5.3.6]). As for classical satellites, an emergency or "survival" mode is defined, the major objective being to ensure sufficient power generation and provide a communication link with the mission control center. To support this mode, the MSU has access to two dedicated Sun sensors.

For position estimation, two nine-channel GPS receivers are available, and for monitoring the maneuvers three two-axis accelerometers.

When the local link is established, the GPS measurements from the receiver of the ISS are available within the GNC of the ATV. Selecting the measurements of the same GPS satellites as taken

by the ATV receiver allows cancellation of common mode errors and provides high-precision "relative" navigation with respect to the ISS (sometimes called relative GPS). But in the close vicinity of the ISS even this accuracy is not good enough. Moreover, RF reflections from the ISS may disturb the measurements (multipath effects). Therefore, for distances smaller than about 200 m a video camera is used for relative position and bearing angle measurements. During the last part of the approach (< 40 m) relative attitude measurements can be derived by evaluating the image of the target pattern mounted close to the docking port.

For independent monitoring of this system two laser-based telegoniometers are available. Like the camera, these sensors provide range and bearing angle measurements.

The attitude control is performed fully autonomously. The required torques are provided by short impulses of the attitude control thrusters. To reduce fuel consumption during the long-lasting drift phases, a specific control mode is activated which allows one thruster only to fire within a control cycle.

During the mission, large transfer maneuvers and small correction maneuvers are distinguished. The first are performed with the four main engines of the orbital control system. They require a proper alignment of the ATV to obtain the desired thrust direction; generally an appropriate slewing maneuver is required before and after. The smaller correction maneuvers can be performed with attitude control thrusters. Their configuration allows thrust to be generated in all three axes so that appropriate thruster commanding gives a thrust vector in any direction, thus avoiding the need for a dedicated slew maneuver.

The control of the 28 ACTs is not trivial. The GNC computes from the deviation between the set point and the actual state a continuously varying request of forces (from position control) and torques (from attitude control). This force–torque vector needs to be translated into appropriate commanding of the thrusters. This is done within the thruster management function (TMF). Since the thrusters are either on or off, neither the force nor the torque is used but rather the impulse. The maximum impulse corresponds to the thrust level times the control cycle time of the discrete controller. Modulating the time gives a quasi-linear variation of the impulse. To save fuel the number of thrusters involved to realize the required impulse is minimized; that is, there is a linear optimization problem, as given below:

$$\begin{bmatrix} f_{1x} & f_{2x} & \cdots & f_{nx} \\ f_{1y} & f_{2y} & \cdots & f_{ny} \\ f_{1z} & f_{2z} & \cdots & f_{nz} \\ t_{1x} & t_{2x} & \cdots & t_{nx} \\ t_{1y} & t_{2y} & \cdots & t_{ny} \\ t_{1z} & t_{2z} & \cdots & t_{nz} \end{bmatrix} \cdot \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \alpha \cdot \begin{bmatrix} F_x \\ F_y \\ F_z \\ T_x \\ T_y \\ T_z \end{bmatrix}$$

with $|u_i| \leq 1$ and $0 \leq \alpha \leq 1$

or

$\underline{A} \cdot \underline{u} = \alpha \cdot \underline{C}$,

It can be solved by the well-known simplex algorithm (see [5.3.8]). The u_i components give the percentage time of the control cycle needed to realize the requested impulse. Multiplying this by the control sample time gives the command to be issued to the thrusters (more precisely, to the PDE units).

This approach is rather flexible, particularly in view of configuration changes as a function of mission success or failure. However, due to the iterative search for an optimum solution, convergence within a predefined number of iterations cannot be guaranteed; that is, deterministic behavior with respect to time as needed for a real-time system is not ensured. On the other hand, simple termination of the optimization process after a predefined number of cycles does not guarantee an acceptable solution.

An alternative to this optimal approach is a suboptimal approach. In this concept a fixed allocation of thrusters to a well-defined force–torque impulse is stored in terms of tables. Since these tables are valid only for a fixed set of thrusters, they need to be defined for each possible configuration, that is each failure case triggers the need for a new table. In addition the solution is suboptimal and less accurate. Nevertheless, the differences are small and it provides deterministic behavior; therefore it was the preferred solution for the ATV.

5.3.6 Verification and Test

The space environment constraints (vacuum, zero *g*, temperature, illumination, etc.) are difficult to mimic on the ground. As a consequence, **simulation** by appropriate simulators [5.3.4] is needed for the verification process. The driving element is the correct mathematical modeling of the environment and all perturbation effects on the equations of motion, including disturbances from the vehicle system itself (e.g., fuel sloshing, flexible appendices). Also the stimuli of sensors (e.g., stimulation of star sensors with a star pattern, or the stimulation of a GPS receiver with an appropriate RF signal) need to be simulated.

The orbital mechanics including perturbations like air drag, nonspherical gravitational field, Earth's magnetic field, solar activity, influence of the ionosphere on RF signals, etc., are well known. Flexible appendices and fuel sloshing can also be modeled with high accuracy. Compromises, however, have to made with respect to model complexity and fidelity (i.e., the model has to be simplified to obtain reasonable computation times for real-time testing, or simply to reduce cost). In the development phase the effort and cost of the implementation and verification of the test environment are quite significant and may well exceed the cost for the development of the item to be tested. However, since these are on the ground they can be derived or adapted from preceding projects or provided and reused in successive projects if developed according to an appropriate standard (ECSS for space).

Nevertheless there are always uncertainties within the defined parameters and the impact of neglected effects causes deviations when confronted with reality.

The ultimate test is the real flight, or a dedicated **demonstration or test flight**. Due to the high related costs, a demonstration flight of the ATV was not planned, but a number of specific tests were embedded in the first ATV (*Jules Verne*) mission to increase confidence in mission success. This included the execution of a CAM maneuver well away from the station, a mission interruption (within the KOS), and a retreat to S_3, where an appropriate escape maneuver was commanded from the ground followed by the initiation of a new sequence of phasing, proximity operations and a new rendezvous approach. Also the second RVD approach was stopped 20 m in front of the docking port and a withdrawal was commanded. Only the third approach was completed and ended with a successful and very accurate docking.

Bibliography

[5.3.1] Fehse, W. *Automated Rendezvous and Docking of Spacecraft*, Cambridge Aerospace Series, Camridge: Cambridge University Press, 2003.

[5.3.2] Interface Definition Document for International Space Station Visiting Vehicles. *SSP 50235*, January 25, 1999.

[5.3.3] Mueller, D. Relative Motion in the Docking Phase of Orbital Rendezvous. *AMRL-TDR-62-124*, 1962.

[5.3.4] Sommer, J. The Functional Engineering Simulator for the ATV RVD Mission with the ISS. ESA Workshop SESP, 1998.

[5.3.5] Mosenkis, R. ATV Executive Summary. *ATV-AS-PRE-1057*, 2003.

[5.3.6] Wertz, J.R. *Spacecraft Attitude Determination and Control.* Dordrecht: Reidel, 1986.

[5.3.7] Junkins, J.L., Turner; J.D. *Optimal Spacecraft Rotational Maneuver.* Amsterdam: Elsevier, 1984.

[5.3.8] Numerical Recipes. http://www.library.cornell.edu/nr/bookcpdf/c10-8.pdf, June 2007.


KRUPP
KRUPP

6 Mission Operations

Thomas Kuch

The mission operations phase after liftoff of the launch rocket shows whether all previous considerations have proven to be correct and will lead to a successful mission. Public interest in a mission – which shows the fascination of space flight – is especially high during this early phase and at critical moments of a mission. Then mission operations are in full focus and often determine mission success or failure.

The objective to explore new, unknown paths constantly encourages new ideas from scientists, engineers and technicians. Therefore, no one mission is the same as another. Suspense will always be immanent in future missions even though it is the ultimate goal of mission preparation to define detailed mission scenarios and eliminate flaws early before liftoff. For this reason the preparation of a mission is initiated many years in advance of the spacecraft launch, depending on the complexity of the mission.

A significant part of the operations of the space segment is conducted from the ground. Thus the correlation between hardware, software and involved personnel is nowhere more visible than in mission operations. It is here where the operations infrastructure, which comprises numerous antennas, computer systems, communication links and control rooms, is defined and subsequently designed, and the ground operations system developed and integrated. Particularly the interfaces between the individual components of the ground segment, but also most notably the space-to-ground interfaces, require the utmost attention. Not least, the definition of operations processes and the writing of procedures requires a considerable amount of the preparation phase when all conceivable failure cases and corrective measures have to be identified and translated into procedures. Because it is impossible to cover all cases of anomalies in advance, the reaction of the operations teams to unforeseen events has to be defined so that the best possible solution can be rapidly determined.

After implementation of the ground operations system, an intensive test and validation phase is initiated. Computer and communication systems, operations procedures as well as the personnel have to be tested and certified. Step by step the risk can be reduced to an acceptable level by conducting additional training sessions and numerous simulations.

In the following sections the complex task of initiating, monitoring and controlling mission operations will be described, depending on the mission type. The relevant infrastructure, control center and ground station network play a significant role and are responsible for much of the effort and reliability of a mission. Using the example of European participation in the International Space Station (ISS), the last section of this chapter describes the special features of mission operations for human space flight.

6.1 Spacecraft Operations

Manfred Warhaut

Spacecraft operations (platform and payload) require intensive long-term preparations. Several years prior to launch the flight control team build-up is started

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9

and the development and configuration of the ground segment and operations system are initiated.

6.1.1 General Concepts and Principles

In general, mission preparation, planning and execution are independent of the mission type. All operations are conducted according to procedures laid down in the **flight operations plan**, a comprehensive document prepared by the flight control team based on project/industry deliverables (user manual and database), the science operations plan and agreements with the principal investigators (PIs). Payload operations specifically are based on experiment user manuals, procedures and databases which the PI teams are required to produce and deliver to the prime contractor, who will integrate them into a single spacecraft user manual.

Spacecraft operations during all active mission phases are generally carried out following an **"off-line" approach**. All activities are preplanned and the resulting telecommands are uplinked to the spacecraft for time-tagged execution (on-board mission timeline). Telemetry evaluation is also mainly off-line, with the limited possibility of quasi-real-time intervention in selected critical phases and in major contingency cases.

The **radio frequency contacts** between the operations control center and the spacecraft are in most cases not continuous and are primarily used for preprogramming autonomous operation functions on the spacecraft, and for data collection for subsequent off-line status assessment. The downlink will normally be so configured that most of the bandwidth is dedicated to the dump of telemetry stored on-board, with limited housekeeping and/or event telemetry transmitted in real time during the pass according to spacecraft monitoring requirements. Anomalies will usually only be detected on the ground after a delay which as a minimum corresponds to the travel time of light, but typically will rather be in the order of one to several hours.

The PI institutes which have developed the scientific instruments will also define their on-board operations. The primary responsibility for developing the payload operations strategy for the scientific mission lies with the complete science working team.

Payload science operations are already defined prelaunch according to user requirements, considering constraints from both spacecraft and ground segment systems. Planned scenarios are validated (prelaunch) by the spacecraft prime contractor and checked by the science operations center. Some mission phases are planned without ground support for payload operation (i.e., the instruments will remain inactive or in basic, safe and autonomous low-activity modes).

Operations during the routine operations phase are preplanned and spacecraft health and mission progress monitoring is conducted off-line after delayed recovery of the required telemetry. This implies a high **level of autonomy** to be implemented on-board the spacecraft at system, subsystem and instrument levels.

In case of failures, ground intervention will normally take place a considerable time – depending on the mission phase – after the intervention of the on-board systems. The approach on the ground will be to collect the necessary information on the failure and the actions already taken autonomously on-board and, if necessary, to reconfigure the spacecraft subsystems or instruments to reestablish the generation of mission products.

Operations (for both spacecraft and scientific payloads) are only conducted in strict compliance with validated event sequences and procedures documented in the flight operations plan. This comprises all operations activities, namely special operations and contingency operations, as well as routine operations for the mission operations phase. Changes in procedures and timelines produced after the launch are managed through the configuration control procedures applicable to the flight operations plan. This includes approval signatures from all parties involved; for timelines the signatures of the representatives of the project office and the spacecraft operations manager are required.

The science telemetry packets are in most cases not processed at the control center. Any information that the control center needs to carry out nominal and contingency operations for the payload must be included either in the nonscience telemetry housekeeping packet or in event telemetry packets. The

precise definition of all the **telemetry packet types** is contained in a specific space-to-ground interface control document.

The control center retains the right to switch off any experiment deemed to be interfering with or endangering the mission objectives, using agreed and validated contingency procedures.

Upon completion of each major mission phase (see Figure 6.1.2), the flight control team compiles a mission operations report summarizing the activities undertaken.

A typical ground segment for spacecraft operations (Venus Express/ESOC) is depicted in Figure 6.1.1.

6.1.2 Mission Types

Depending on the mission objectives, different orbits or flight trajectories are selected to fulfill the requirements for data acquisition, transmission, processing and distribution in the most efficient way.

6.1.2.1 Spacecraft in Low Earth Orbits

Low Earth orbit missions are typically characterized by an orbital period of 90 minutes and brief ground station contact periods of about 10 minutes. This requires a high degree of **automation** to configure both the spacecraft and the ground station for data transmission to ground. The payload data volume is often very large (megabytes to gigabytes), as in the data acquisition and **transmission of pictures and images**, for example.

6.1.2.2 Spacecraft in Highly Elliptical Orbits

Highly elliptical orbits with durations of 12, 24, 48 or 72 hours are primarily used for spacecraft designed to observe the Milky Way and/or other galaxies at different

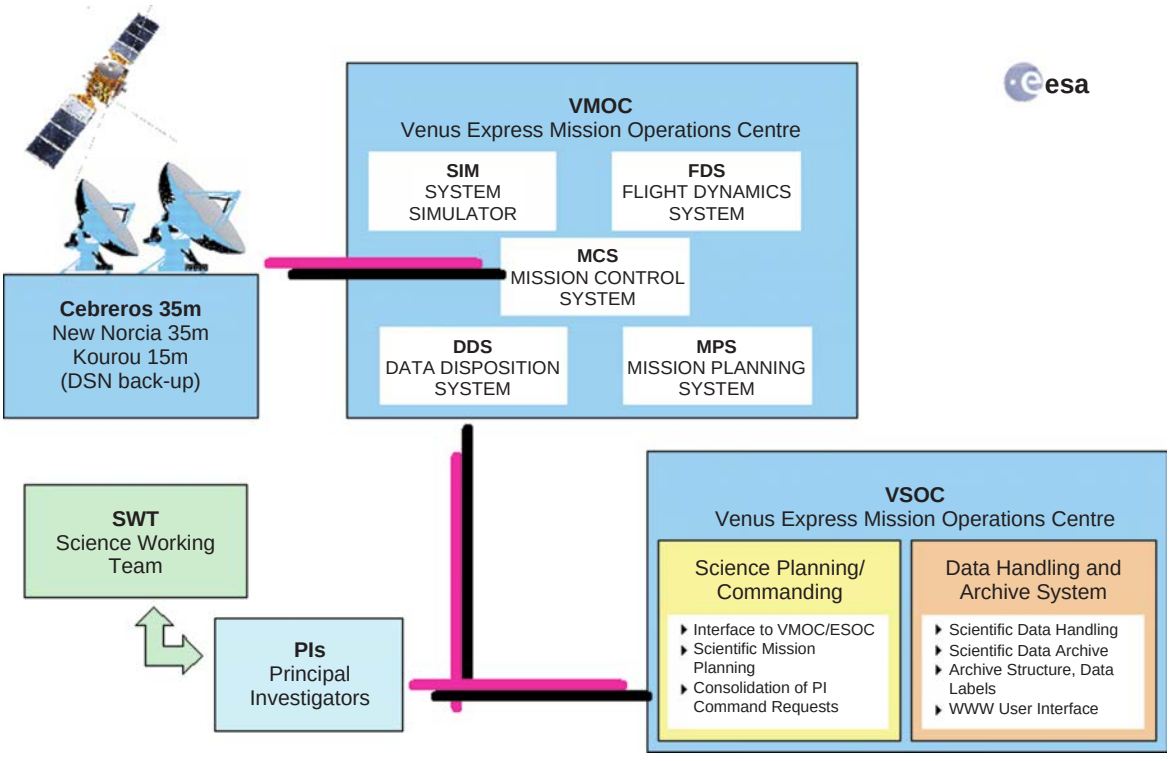


Figure 6.1.1: *Ground segment for Venus Express (ESOC) (Source: ESA).*

spectral wavelengths (X-ray, gamma ray, ultraviolet or infrared). Utilizing two ground stations at selected longitudes allows ground station coverage of about 14 to 16 hours. The data rates for transmission to ground are typically kilobits to megabits per second and allow contact periods long enough to also conduct spacecraft operations in real time.

6.1.2.3 Geostationary Spacecraft

The geostationary orbit is almost exclusively used by **telecommunications and meteorological** spacecraft. The quasi-stationary position over a predetermined region allows continuous transmission of radio and television programs as well as images in the optical and infrared ranges for meteorological applications.

6.1.2.4 Spacecraft at Lagrange Points L1/L2

Nowadays, astronomy and observatory missions are injected into an **orbit around Lagrange point** L1 (toward the Sun) or L2 (away from the Sun) at a distance of 1 million or 1.5 million kilometers, respectively, from Earth. These orbits are characterized by small temperature variations and low radiation exposure. They avoid passages through the Van Allen radiation belt. Contact periods with the ground stations, the data transmission rates and the mission operations concepts for these spacecraft are similar to those used for satellites in highly elliptical orbits.

6.1.2.5 Interplanetary Spacecraft

Interplanetary missions are serving to explore planets, moons, asteroids and comets in our Solar System or to observe the Sun. The ground contact periods of 8 to 12 hours are mostly used for data acquisition in **playback mode** and to configure new, subsequent on-board timelines. Data transmission rates are low (bits to kilobits per second) and the signal reaches Earth only after several hours. Moreover, due to spacecraft–Earth–Sun geometrical constraints, radio-frequency signal transmission might be lost during the mission, causing so-called communication blackout periods of several days or even weeks.

6.1.3 Pre-Mission Support and Implementation Activities

Based on the scientists' ideas and concepts, the spacecraft's operations tasks and requirements are considered at the start when a space mission is defined. In these early project phases the feasibility of a project is analyzed and determined in both technical and scientific terms. During the study phase (**assessment study**), which usually occurs several years prior to the planned launch, mission requirements and constraints are analyzed according to the following considerations:

- Selection of orbit or trajectory
- Spacecraft configuration
- Payload and spacecraft subsystems
- Launch and mission design
- Ground segment infrastructure
- Mission operations concepts.

At the end of the study period, appropriate results and possible alternative solutions are documented in a study report.

The subsequent **Phase A study** comprises generic requirements analyses for mission control and ground segment infrastructure for all mission phases. The mission phases are divided into a launch and early orbit phase (LEOP), a commissioning and in-orbit test phase, routine mission operations and a run-down phase. The Phase A study covers amongst other things data acquisition, transmission, processing and distribution, navigation and flight dynamics, and ground station and communication network aspects. Moreover, a coherent concept for maintenance and operation of the flight and ground software and ground segment elements is established. The previously established study report is refined, detailed and amended, together with a relevant cost estimate.

Formally approved projects are typically implemented in the subsequent project phases as follows (see also Chapter 2):

- Phase B: Definition
- Phase C: Design, development and implementation
- Phase D: Test, training and simulation

- Phase E: Launch and operations
- Phase F: Run-down phase.

The responsibility for all ground segment and mission operations activities is delegated to a ground segment manager acting also as flight operations director during the mission-critical phases. Responsibility for the routine mission operations is vested with the spacecraft operations manager leading the flight control team.

A typical **mission evolution** for a science project is depicted in Figure 6.1.2.

6.1.3.1 Mission Analysis and Systems Studies

In the context of spacecraft operation activities the following mission analysis and systems studies are carried out:

- Launch window analysis
- Evaluation of **geocentric and interplanetary trajectory** requirements and constraints, including eclipses, ground visibility, maneuvers, payload operations and their relative timelines
- **Mission profiles** and establishment of "delta-v" requirements, including the definition of maneuver strategies
- Alternative operational orbits or trajectories
- Orbit and attitude determination **covariance analysis**
- Ground station coverage
- Inputs/constraints to **science operations planning**.

The results are documented in the consolidated report on mission analysis and in respective study notes. In

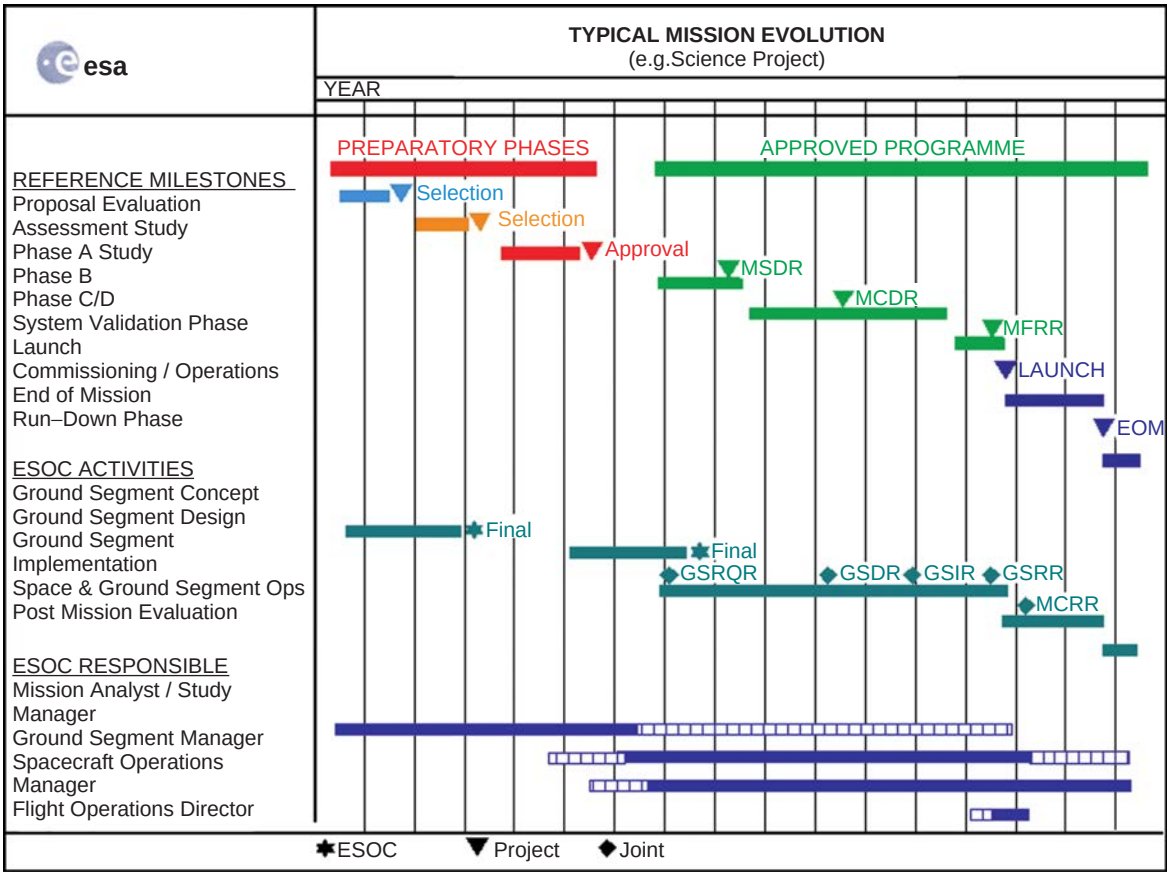


Figure 6.1.2: *Typical mission evolution from the mission operations perspective (Source: ESA).*

addition, required negotiations with the launcher authority are supported.

6.1.3.2 Engineering Support and Implementation Activities

The implementation of mission requirements encompasses mission analysis, ground segment management, project control, flight dynamics, data processing, simulation, communication, computer hardware implementation, network operations, mission control, and station engineering support activities with appropriate engineering support as laid down in the **mission implementation plan**. This includes all ground segment system definition, planning and implementation activities required for the mission as well as the preparation and control of user requirements, preparation and management of procurement contracts for ground system facilities, ground segment design and development, ground segment testing, the definition of a mission control concept, and all mission control preparations. Furthermore, studies and analyses in the various ground segment areas (e.g., mission analysis, flight dynamics, operational feasibility, etc.) are conducted in order to demonstrate the feasibility of the mission.

Operational inputs to the spacecraft design are provided by means of specific implementation documentation and participation in major project reviews.

6.1.3.3 Ground System Validation and Testing

The ground system test and validation activities typically begin around two years before launch, with the exception of some specific tests which are conducted earlier. Activities include tests to confirm radiofrequency compatibility between the spacecraft and the ground segment and to validate flight hardware compatibility with the control center software, the operational databases and flight operations documentation. System testing includes simulation of propagation delays and their testing. The different elements of the ground segment undergo acceptance tests before they are linked together to finally constitute the network. The system simulator developed by the control center is used for mission control system software verification and for personnel training.

Radiofrequency Compatibility Tests

To confirm **radiofrequency compatibility** of the **spacecraft** with the **ground station**(s) an appropriate test is conducted about two years prior to launch. For this purpose the project provides a radiofrequency suitcase model comprising flight-representative hardware (e.g., engineering model units) sufficient to test all uplinks and downlinks for both functional and performance characteristics. This includes verification of all telemetry, telecommand and ranging functions and their combinations, as well as spectral analyses and link budget verification. The test is conducted with a reference ground station representative of all ground stations used throughout the various mission phases. To ensure compliance with ground stations supplied by other partners, similar tests are conducted as required.

Control Center Software and On-board Software Validation

The responsibility for development and validation of on-board software (spacecraft and experiments) rests with the supplier, or the prime contractor at system level. However, the control center handles postlaunch maintenance of the on-board software for long-duration missions. This requires delivery of the complete on-board software together with the related software development environment to be used at the control center for generating software patches. About 12 months prior to launch, special mission education workshops are held to transfer the knowledge from the software specialists to the flight control team in the control center. At that time also the **software development environment** is integrated into the control center and the on-board **software maintenance** capability is validated. Any future software deliveries are subject to strict configuration control and include certification of proper validation and formal acceptance by the project. The prime contractor supports on-board software maintenance at the control center during mission-critical activities after launch, but during routine operations phases only on a call-out basis. The development, testing and validation of the experiment and instrument software are the responsibility of the instrument suppliers or scientists.

Mission Sequence Tests
Mission sequence tests are conducted to validate the feasibility of selected mission scenarios. Based on a typical mission slice comprising representative spacecraft and payload operations, related command sequences for the spacecraft subsystems and the payload are defined, uploaded on-board and executed together with typical ground segment activities. The mission sequence test is performed as soon as possible in the overall project schedule as part of the **spacecraft/payload functional tests** and/or is conducted during integrated satellite tests under thermal vacuum conditions. The test duration is approximately one day for each test and the flight control team supports these tests with the definition of representative mission scenarios.

System Validation Tests
The project provides for online access to the spacecraft flight model for closed loop testing (system validation test) with the ground segment and the flight control software. These tests comprise spacecraft **commanding** from the control center, **data flow tests** between satellite and control center, and real-time nonscience **telemetry data processing** in the control center in parallel with the telemetry processing performed in the checkout equipment. The connection of the control center with the spacecraft is established by means of leased data lines or by the use of existing communications facilities via a network data interface unit presenting the same interface to the control center as actually used by the operational ground station(s). The telemetry and telecommand traffic is visible to the assembly, integration and validation team at the spacecraft integration site.

A second system validation test involving the **spacecraft, control center, mission control software and spacecraft controller** is performed about nine months prior to launch. It includes confidence testing of communication links and final on-board software deliveries. Around two to three months prior to launch a third system validation test is performed for final verification of the integrated ground segment interfaces to the spacecraft, with the spacecraft being located at the launch site. The results serve as inputs to the ground segment readiness review. The final readiness decision is taken in the course of the launch readiness review. The project and the prime contractor review/approve the test procedures and ensure specialist participation (project personnel, prime/subcontractors) at the control center, as required.

System Operations Validation
The system operations validation program aims to execute a series of **end-to-end operational scenarios** to verify the readiness of the ground segment as a whole to support the mission. As such, a number of standard and mission-unique tests are executed, such as an end-to-end commanding test, routine mission operations test, routine phase ground segment contingencies, and LEOP scenario load test and contingencies.

In addition, a set of **mission-unique tests** is planned. A test is performed to validate the capability of recovery of the spacecraft from safe mode, according to procedures supplied by the spacecraft prime contractor. Interface tests for data access by remote users are performed to demonstrate compatibility in terms of physical/logical connectivity and application interfaces (file request/transfer mechanisms, command request capability). This test is performed around 10 months prior to launch, applying an "operational" scenario with multiple users and may include measurements of the turnaround times. This test is repeated about six months prior to launch with the equipment of the PI teams installed at the control center. In case the distribution of data is required on a hard data medium (CD/DVD), a special test verifies the production, data formats and structures. The objective of the control center/user center interface test is to verify the interface functions and procedures required to generate a consolidated operation request schedule ready for subsequent uplink to the spacecraft. Furthermore, all operational interfaces with users or user centers as defined in the interface control documents are exercised. In case the utilization of (ESA-)external ground stations is required, their related interfaces and functions are validated. This includes telemetry, telecommand and tracking interfaces, exchange of state vectors and orbital data for all relevant mission phases.

6.1.3.4 Training and Simulations/Mission Readiness Tests

A **system simulator** which is representative of nominal spacecraft behavior and a number of selected

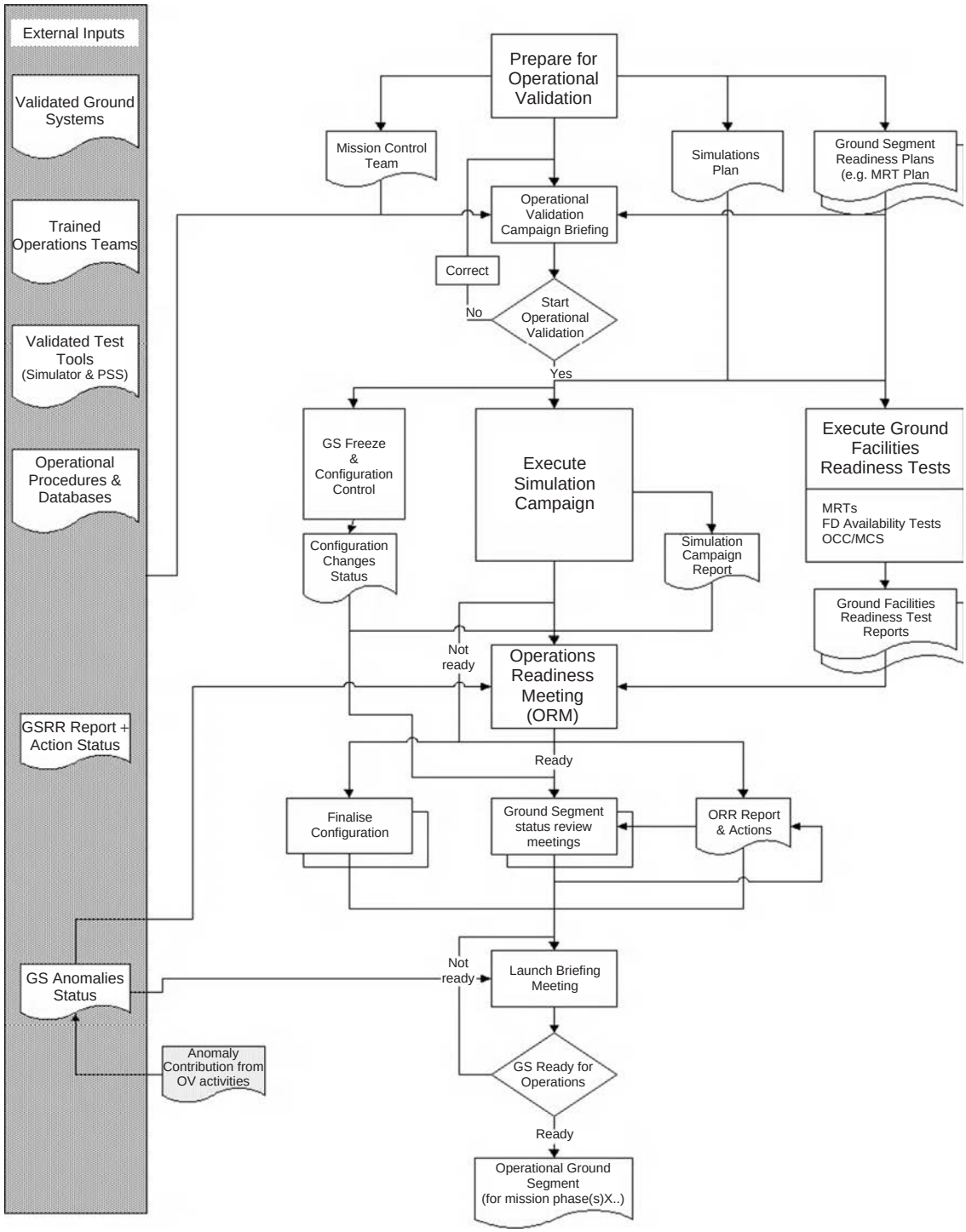


Figure 6.1.3: *Operational validation scheme (Source: ESA).*

failure cases is developed by the control center.The simulator is used for several purposes: as a test tool for the development of **operational software** (telemetry processing and telecommand) and the setup of operational databases; for training operations personnel to the highest degree possible, including the simulations of nonnominal behavior of the satellite and of failure cases, as well as for validating operational procedures. The system simulator is the prime data source for ground segment validation testing, for **training the staff** and for exercising the complete ground system in a predefined series of simulations prior to launch. The system simulator also is required as a means of **validating operational procedures**. As described, the simulator is used prior to launch and throughout the entire postlaunch mission period. The simulations campaign focuses primarily on LEOP and commissioning activities, as well as on critical orbit or trajectory correction maneuvers, whereby some of the activities may be repeated in postlaunch simulations (e.g., preparations for planet fly-bys and planet orbit insertion and correction maneuvers).

The mission readiness tests are designed to prove the readiness of the overall ground station network to support the mission in primary and backup modes and to validate the operational status of the ground system and communications interfaces. These tests are normally extended data flow tests, with additional processing, display requirements and/or specific performance checks. The mission readiness tests are organized and conducted around two months prior to launch.

Spacecraft prime contractor staff and payload specialists participate in the training and simulations program prior to launch, as well as in the actual LEOP activities. For routine postlaunch operations, industrial and/or specialist availability is reduced to anomaly investigation support on a call-out basis, as required. The operational validation flow is depicted in Figure 6.1.3.

6.1.4 Mission Phases

6.1.4.1 Launch and Early Orbit Phase

The LEOP operations typically last from three to seven days, starting from a few hours before launch and ending with the start of the in-orbit commissioning phase. Launch support starts eight hours before the launch and includes a final readiness test with the ground stations. A few hours prior to launch, data flow tests and data confidence tests are performed with the data being transmitted from the satellite at the launch site to the control center. This permits final verification of ground system and network readiness for the final go/no-go decision about 30 minutes prior to liftoff. After spacecraft separation from the launch vehicle (typically 20 to 90 minutes after launch), a series of configuration activities is performed automatically by the spacecraft, including priming of the **reaction control subsystem**, acquisition of a Sun-pointing attitude, deployment of the **solar array** panels and establishment of **radiofrequency contact** with the ground stations. At that time the control center takes over control of the spacecraft and completes the initial configuration activities including configuration of the main subsystems (**data management system, attitude and orbit control system**) and activation of the most important system functions. In parallel, precise orbit determination is established by analysis of the radiometric measurements, and the parameters for the first trajectory correction maneuver are calculated if required.

The LEOP operations are primarily carried out from the **main control room** accommodating the flight director, the project representative, the spacecraft operations manager and flight control team, the ground operations manager and the data systems manager. For missions operated from ESOC the ESTRACK (ESA tracking stations) Control Centre is in charge of the control of the ground stations and the communication network and is manned with a team having multiple-mission responsibilities. In the flight dynamics room a team of specialists is in charge of all activities related to the analysis of the dynamic behavior of the spacecraft and of the calculation and command generation for the execution of orbital and attitude maneuvers. The project support room hosts the project support team, composed of spacecraft experts from the project office and the spacecraft manufacturers, who can communicate directly with the flight control team during critical phases or in case of anomalies. All control rooms have standard consoles with voice loops, alphanumeric and graphical displays of telemetry data for commanding, as well as displays of the actual position of the spacecraft.

During LEOP ground station support is aimed at providing continuous coverage of the spacecraft (geostationary orbit missions) or frequent, regular **contact periods** (low Earth orbit missions). All ground segment elements are often provided with full redundancy, and communications circuits to the control center are routed via diversified data lines. The mission control team in this phase is structured in a traditional way under the supervision of the flight operations director. The **organizational structure** of the mission control team typically used at ESOC is depicted in Figure 6.1.4. The team covers 24 hours of daily operations in two shifts of about 12 hours each.

6.1.4.2 Commissioning and In-Orbit Test Phase

The commissioning and in-orbit test phase starts after completion of the LEOP activities. In this mission phase additional orbit/trajectory and/or attitude corrections are implemented, if required. This mission phase is primarily dedicated to activation and checkout of the **payload**, instrument **calibration** and verification that the overall system performance meets the requirements. This involves amongst other activities a checkout of the payload interfaces with the spacecraft subsystems, and interference measurements between the payloads and the subsystems in order to detect possible influences on the payloads. All subsystems not immediately tested during LEOP operations are also commissioned at the beginning of this phase.

Communications with the spacecraft during this phase utilize **ground stations** allowing high data transmission rates. Experiment specialist participation is required at the control center for the first experiment switch-on operations and subsequent performance checkout. The mission products and the facilities provided to payload specialists and scientists at the control center are defined in Section 6.1.5.4 below.

All **payload instruments** are operated in various operational and calibration modes, followed by an analysis of output measurement data including any relevant instrument or spacecraft housekeeping data. The scientists/users will verify the correct functioning of the instruments using this data. Critical operations such as power-up sequences during commissioning will initially be performed only during ground coverage periods when real-time control and monitoring of the spacecraft are possible. Details pertaining to payload commissioning activities are defined in the flight operations plan.

The **control center operations** are often carried out from a dedicated control room planned for routine operations at a later stage. It consists of a small number of spacecraft control workstations and other workstations (typically 3–4 consoles) for monitoring network and station activities and accessing all off-line auxiliary systems (e.g., data disposition system, database system, simulator, etc.). All operations requiring direct interaction with the spacecraft are conducted from this room.

A special room or area hosts the scientist/instrument teams and their dedicated equipment for monitoring the performance of their instruments. These teams have access to their data streams and are able to submit command requests via electronic links for processing and uplinking.

During the commissioning and in-orbit test phase very often only one ground station is used, which is identical to the one planned for later routine operations support. At the start of the commissioning phase the responsibility for conducting all mission operations is delegated to the spacecraft operations manager and team, namely the flight control team. Network and station operations are carried out by multiple-mission support teams in their respective control rooms. Specialist support for flight dynamics activities and maintenance of generic control center and mission-dedicated hardware and software are provided during normal working hours, but for the investigation of anomalies a call-out scheme applies. The flight control team shift pattern is aimed to coincide with the varying ground station visibility periods. The commissioning strategy is based on an ever-increasing use of the on-board mission timeline with the resulting telemetry obtained in near real time. The PIs are then expected to follow the same working hours, in order to correct or prevent anomalies without losing a single day.

6.1.4.3 Routine Operations Phase

The routine operations phase duration is adapted to the mission needs and initially lasts just two years, but

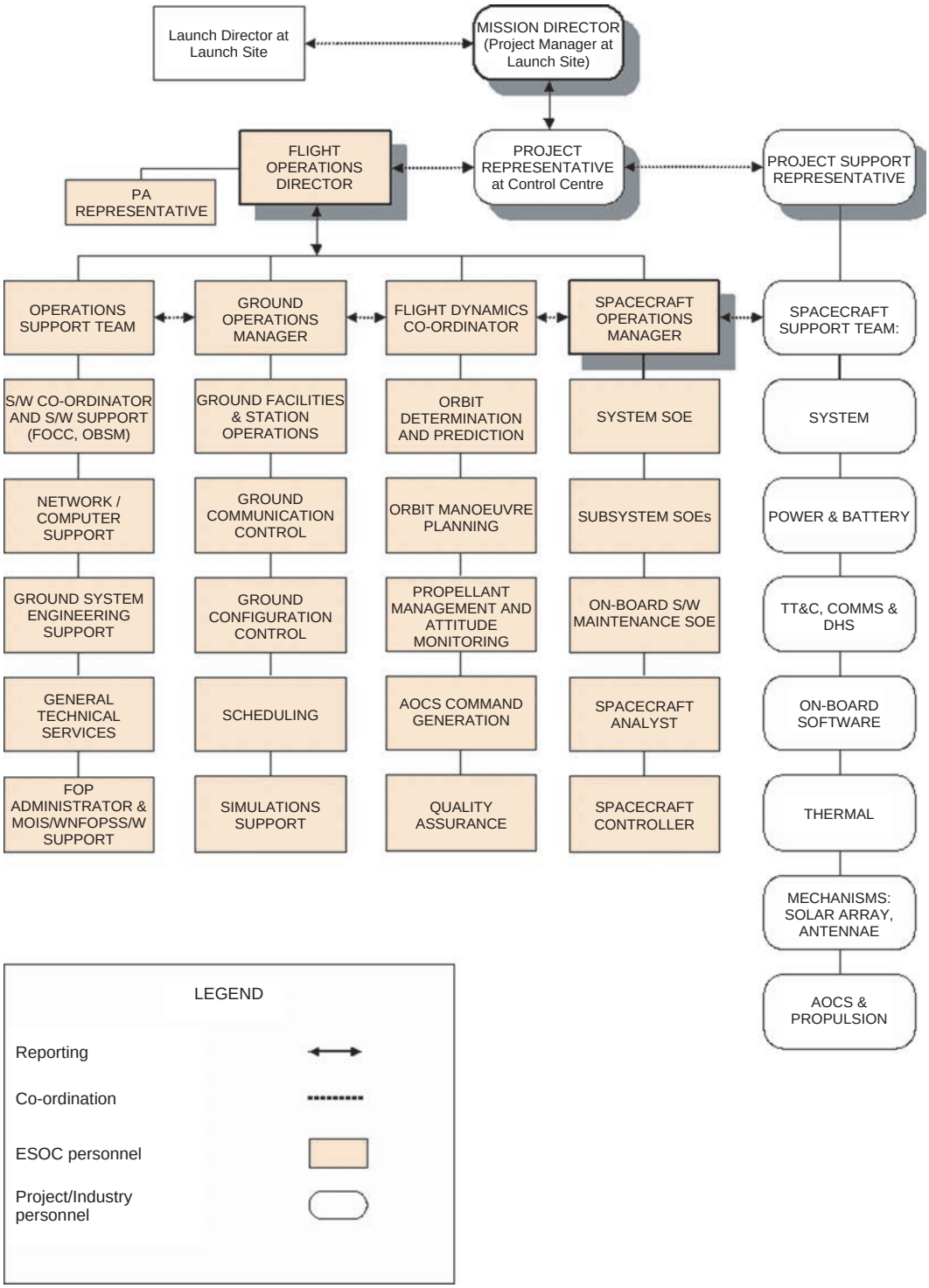


Figure 6.1.4: *Typical LEOP mission control team at ESOC (Source: ESA).*

it could be extended up to about 15 years if desired, possible and funded. In this phase various **science objectives** are pursued according to the **master science plan** established by the scientists and user groups. This plan provides the complete set of guidelines for science operations as well as the selection of orbit and attitude strategies. The mission planning cycle further details the plan and translates it into operations timelines using special mission planning software. Routine mission operations are carried out from a dedicated control room. Station and network control are carried out from a multimission control room.

The **mission control team** staffing remains at the reduced level reached in the previous mission phase and the team conducts operations on a daily basis working a normal shift pattern. Support personnel carry out network and station operations. There is no requirement for scientist/expert support to be present at the control center. Support for the investigation of anomalies can be ensured through a call-out scheme as necessary. For major anomalies or specific critical/contingency situations, expert support is available through related contractual provisions. Requirements for flight dynamics support and for maintenance of generic control center systems are identical to those for the commissioning and in-orbit test phase.

6.1.4.4 Recovery Operations

In case a major launcher or spacecraft **contingency** makes it impossible to achieve the mission objectives on the basis of the available mission timeline and procedures, the original mission is terminated and if possible redefined. In this special **emergency** case the mission control team develops contingency procedures under the lead of the flight director and mission manager. If possible a new mission will be defined, whereby additional facilities and/or new contingency procedures will be specified and implemented. A revised flight operations plan, comprising new or updated procedures on the basis of actual spacecraft status, will be used to operate the new mission. This will, in particular, take the spacecraft's health and safety into account in order to ensure the success of the redefined mission.

6.1.4.5 Run-down Phase and Deorbiting

Spacecraft **lifetime** is often limited by available on-board consumables (fuel, batteries) and the lifetime of on-board transmitters/receivers. Taking into account the increasing problem of **space debris,** policies are being developed to require a controlled run-down and deorbiting of spacecraft in low Earth and geostationary orbits. Spacecraft close to Earth are reentered into the atmosphere in a controlled manner in order to burn up the vehicles as completely as possible. Geostationary spacecraft are maneuvered into a **graveyard** orbit (somewhat higher than the geostationary position) to free their location for new ones. The run-down phase operations include the shutdown of instruments. Thereafter, the flight control team archives software and documentation, and compiles a "postmission evaluation report" highlighting the major mission events and addressing the lessons learned during the course of the mission, as well as documenting the final cost at mission completion.

6.1.5 Mission Operations Tasks

6.1.5.1 Spacecraft Monitoring

Spacecraft **telemetry data** is controlled and monitored on a routine basis. Telemetry data analysis allows determination of trends and possible critical out-of-limit conditions.

6.1.5.2 Anomaly Handling

The flight control team prepares a formal **anomaly report** comprising all operational errors or anomalies of the ground segment, the spacecraft and the payload. Where required the spacecraft operations manager requests detailed investigations and analyses from the prime contractor in order to explain the anomaly in more detail in a separate report also containing proposals for the procedure to close or avoid the anomaly.

Upon receipt of the anomaly report the spacecraft operations manager immediately initiates a **failure analysis**, together with the definition, implementation and validation of appropriate corrective actions. After

completion of the analysis and implementation of the recovery action, the spacecraft operations manager distributes a close-out report. All anomaly reports are stored in a database accessible to all project members to provide them with current information about the actual status of the mission.

6.1.5.3 Spacecraft Operations

Spacecraft operations activities are dependent on the mission phase and the criticality of operations. During ground station **contact periods** the control center receives telemetry data from all spacecraft subsystems and the payload. Data acquired and stored on-board outside the coverage periods is dumped at high data rates to the ground. All **telemetry data** is analyzed upon receipt in detail utilizing mission control system software. This includes calibration and display of selected telemetry parameters, comparison of analog or digital telemetry parameters with preset limits or expected values, triggering audible and visible alarms in case of out-of-limit conditions and the calculation of derived parameters combining different received telemetry data measurements of the spacecraft. This allows a quick check of the summary status of the spacecraft and the possible identification of on-board anomalies.

The flight control team analyzes the received and processed information at the control center and determines the **spacecraft status** on the basis of the telemetry data. The reaction to detected anomalies is based on available contingency procedures. In the case of anomaly situations not covered by existing contingency procedures, specialist support can be called at any time during or outside contact periods with the spacecraft. In addition to the monitoring and control activities, spacecraft operations require regular **commanding** to maintain functionality and to execute the planned operational timelines. Typical examples of spacecraft commanding activities are: spacecraft configuration to enable ground station contacts, like turning on the on-board transmitter and configuring telemetry transmission; starting or stopping data dumps from on-board storage devices; executing attitude maneuvers, maintenance of orbit or correction of trajectory; periodic or seasonal calibration and configuration of sensors; periodic tests of spacecraft subsystems and payloads.

The amount of **manual commanding** initiated by the flight control team during ground station contact periods is limited to **minimize risk** of operational errors. Most spacecraft operations are preplanned and the resulting telecommands are uplinked to the spacecraft for time-tagged execution (on-board mission timeline). Telemetry data transmission is continuous during contact periods and amounts to several minutes for near-Earth missions and several hours for interplanetary missions, or is permanent for geostationary satellites. Commanding activities are, however, always performed over relatively limited time periods.

6.1.5.4 Data Acquisition and Distribution

Data Access

During the entire mission a **data disposition system** at the control center allows for external rapid access to the most recent data available via communication lines on a call-up basis. The data includes telemetry and auxiliary data, as well as related catalogs. Telemetry data (spacecraft and payload) is provided as raw data; that is, time-stamped packets individually stored on logical files according to the application identification (ID) consisting of process identifier and data type. Auxiliary data generated by the control center contains information needed to assist in processing and analyzing the science data, and also to support mission planning and command request generation. The data may be stored on different files according to its nature. The catalogs contain a full record of all data sets available and the time period to which each data set pertains. The data disposition system is a near-real-time processing system which provides data access on a demand-driven basis; that is, the user or delegated representative is responsible for generating appropriate data requests. The data/message transmission is via file transfer.

Data access is possible either at the control center or from a remote location and access to the data is limited to authorized users. Confidentiality agreements might be necessary in addition. The complete data set is kept online for the whole mission duration. File transfer requires network functions and protocols to be compatible with FTP (TCP/IP). The users are responsible for providing the required terminals or

workstations at the remote host site, and the leasing/rental of public communication lines.

A **data delivery interface document** governs pertinent interfaces. This document is issued and agreed with all users. As a formal project document it is maintained under configuration control throughout the duration of the project. It describes the formatting of delivered data down to the level of detail necessary to enable users to retrieve science data and any required housekeeping or auxiliary data.

Command Request Handling
In addition to the data access capability, the system allows for transfer of consolidated **command requests** from the user center to the control center as inputs to the mission planning system during routine operations. The system supports approval, authentication and authorization of command requests. After validation the control center incorporates the command requests into the mission planning system, which generates the final command schedule for uplink to the spacecraft.

Critical Payload Operations
A special control room or area for users and scientists at the control center provides data access and payload command request capabilities, as well as communication with remote locations (e.g., via the Internet). The interface is identical to the remote interface and supports both telemetry delivery services to the experiment's electrical ground support equipment and special command requests from users.

Raw Data Medium Distribution
The distribution to users of all science data and the necessary auxiliary data on a raw data medium (CD, DVD or similar storage medium) is supported as required.

Raw Data Long-Term Archiving
Raw telemetry and auxiliary data is kept at the control center throughout all postlaunch mission phases in a long-term archive accessible through the data disposition system.

Data Delivery Formats
Each data delivery request to the data disposition system results in the transfer of a block of data containing three main areas: an **acknowledgment**, including request details and status; a **catalog entry** giving identification details of the requested data actually supplied (e.g., experiment, date, time); and the requested data itself.

Simple packaging in **standard formatted data units** (SFDUs) will be applied, following a recommendation of the Consultative Committee for Space Data Systems (CCSDS). Apart from providing a convenient mechanism for handling the variable length of requested data, this standard also provides administrative support for the description of application data. Both the formatting of data delivered through the data disposition system and long-term data archiving follow this standard.

Data Recovery
Scientific data acquisition is maximized to be compatible with achievable space-to-ground link margins. The detailed data collection requirements are derived from the payload interface document and the operations rules and constraints document for each experiment. Consistent with this aim, the strategy for scientific operations and scientific data recovery are defined in the pertinent master science plan applicable to the required mission phases (commissioning, routine operations phase).

6.1.5.5 Mission Planning

In a typical mission planning scenario the user teams provide, at fixed deadlines and with a fixed periodicity, inputs to the user center for the requested science operations. The user center passes a consolidated request to the control center, which checks the requests against mission, environmental and resource constraints. For the routine science operations phase, the timeline of the spacecraft attitude and the season (eclipses, occultations, Earth distance, etc.) play a major role in establishing the constraints scenario against which the payload operations plan has to be checked.

6.1.5.6 Navigation and Flight Dynamics

The term **flight dynamics** covers all operations related to spacecraft orbit/trajectory and attitude. The monitoring of spacecraft attitude encompasses analysis of

telemetry data delivered by on-board sensors, gyros, star trackers, Sun sensors and Earth sensors. These allow precise **determination of the attitude** as well as monitoring and control of the proper functioning of the related sensors and elements. Attitude determination includes command generation for the on-board attitude and orbit control system of the spacecraft, in particular its thrusters and control mechanisms, in order to correct the attitude by means of special maneuvers.

The **orbit determination** of a spacecraft is performed through analysis of **radiometric data** obtained from the ground stations. Doppler data (shift of radio-frequency signals) is analyzed during the respective contact periods and the spacecraft orbit/trajectory is continuously recalculated and compared to the expected orbit or trajectory. Depending on the mission type, the spacecraft velocity can be determined to an accuracy of fractions of millimeters per second and the spacecraft position to an order of meters or kilometers. To determine the trajectory of interplanetary spacecrafts which have to fly by bodies in our Solar System like comets and asteroids whose position is not known accurately enough, Doppler measurements and range/range rate measurements are insufficient. Therefore, other methodologies like optical navigation are required. This method uses **images of celestial bodies** obtained by on-board cameras. After analyzing several of these image sequences, the relative position and velocity can be determined to high accuracy.

In case the measured trajectory deviates significantly from the planned trajectory**, correction maneuvers** have to be planned and implemented with on-board thrusters and the propulsion system. Trajectory correction maneuvers are also required at specific times during the mission. After launch of a spacecraft and separation from the launcher, the desired orbit or trajectory has to be reached. For this purpose orbit/trajectory correction maneuvers are required and the flight dynamics team has the task of optimizing them by calculating related parameters like thrust durations and to generate resulting commands for transmission to the spacecraft. The results obtained from **flight dynamics calculations** are required for routine spacecraft operations. They include the prediction of basic mission parameters like the start and stop of ground station contact periods and eclipses, as well as orbital parameters to allow pointing of ground stations towards the spacecraft.

Additional flight dynamics activities include monitoring and control of the special dynamic behavior of the spacecraft, such as boom oscillations, nutation and spin velocity of spin-stabilized spacecraft, the calculation of fuel utilization and remaining fuel, as well as the resulting changes to the center of gravity.

6.2 Control Center

George Hiendlmeier and Thomas Kuch

The text below, in its details, refers to the German Space Operations Centre (GSOC) of the DLR at Oberpfaffenhofen, near Munich. This control center differs from other control centers in that it hosts operations facilities for the support of human space flight missions as well as of unmanned space missions. However, the general principles are valid for other control centers around the world as well.

Computer systems employed inside a control center can be grouped into systems used to maintain external data interfaces, for routing and logging of data, to process data (telemetry and telecommand) and to visualize data. These systems are complemented by systems used for mission planning, flight dynamic calculations and so-called operations tools (e.g., Ops-Web, which allows information exchange via the Internet or Intranet).

6.2.1 Control Rooms

Control rooms are the “business card” of a control center. But how did they come into existence? In the initial days of space flight it was necessary to combine expensive specialized hardware systems in dedicated rooms. Work areas for engineers and scientists were co-located with the hardware, and the data networks used today were unknown. Additional benefits were the easier security control and the existing exploratory and team spirit of the early years.

The growing interest of the public in space flight activities necessitated a public-relations-oriented outfitting of the control rooms. Who does not remember the first pictures from NASA's control rooms at Houston, Texas, during the Apollo missions and the lunar landings?

Today, the hardware used in control rooms mainly stems from the standard lines of different manufacturers. **Networking technologies** allow distributed solutions in different rooms or even disbursed over various control centers. Increased **security requirements** have evolved from the necessity to exclude unauthorized persons from disrupting space operations and/or accessing proprietary data. Requirements toward **ergonomic design** and user-oriented interfaces have become important criteria in the design of control rooms.

The need for a strong team spirit in supporting the cooperation of the whole team during important mission phases is unchanged. The ever-increasing interest of the public and the presence of the media have raised the quality requirements for presenting the activities during ongoing operations.

Different kinds of control rooms are found in a control center (Figure 6.2.1). Mostly shown to the public are the ones hosting mission operations. This is where experts on the different subsystems of a spacecraft are co-located with the flight director and representatives of the owner and/or spacecraft manufacturer during the course of a mission. Depending on the mission goal, the rooms are occupied for weeks or months – in the case of the International Space Station, even for several years.

Mission control rooms in general are furnished with a number of consoles in a standard configuration (Figure 6.2.2 and Figure 6.2.3). The standard configuration contains two to three workstations each with one or two flat screens, a user station for the digital

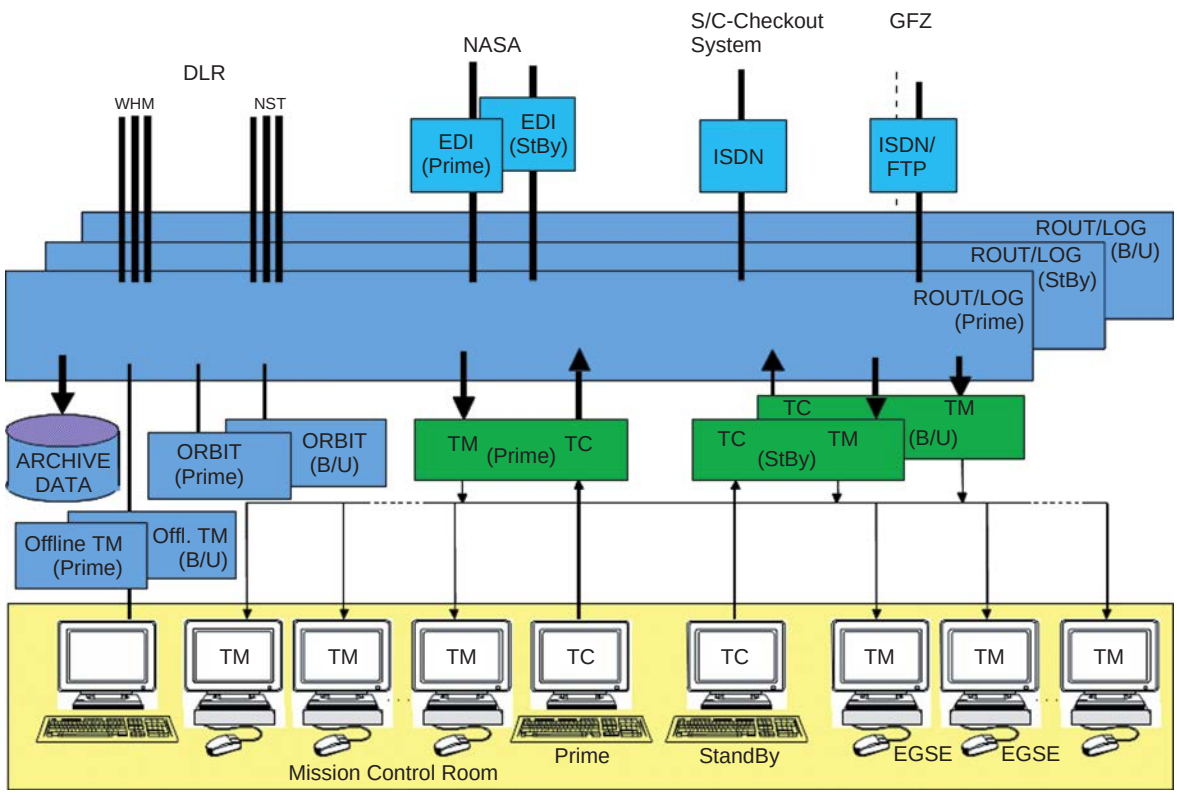


Figure 6.2.1: *Control center computer structure (Locations: WHM = Weilheim, NST = Neustrelitz) (Source: DLR).*



Figure 6.2.2: *Typical mission control room layout (Source: DLR).*



Figure 6.2.3: *Control room at GSOC (Source: DLR).*

voice conferencing system (see Section 6.2.2.2) and a telephone.

In addition, control rooms feature large projection screens and video monitors to display information of general interest.

So-called **back rooms** (Figure 6.2.4) are provided for mission support like ground network control. One typical type is a dedicated room for experts in flight dynamics. Besides specially configured workstations, these rooms provide conference space for discussions within the expert teams. These necessary discussions would disturb operations in a mission control room. Because of their specific know-how, these experts usually support more than one mission at a time.

For similar reasons, special rooms exist for the configuration and handling of computer systems, for the communications network and for ground station coordination.



Figure 6.2.4: *Control room for network operations (Source: DLR).*

6.2.2 Hardware Components

A control center hosts numerous computer systems assigned to different tasks. A medium-size control center like GSOC uses approximately 1000 computers.

6.2.2.1 Computer and Network Architecture

The data network uses standard products available from industrial networking technology sources. In the case of GSOC, this is a centralized (collapsed) **backbone** made from CISCO components. Users and computer systems are connected via fiber links. Central server systems are built from PC-class (Intel) computers of varying capacity. Based on special requirements from supported projects other technologies (e.g., Sun) may be used. The peripheral man–machine interface computers (workstations) use Intel-class computers. These are tailored to the required computing power. Another major selection criterion is the noise emission of the computer.

The **operating systems** have evolved from real-time operating systems (RMX, VMS, Unix) to Linux and MS Windows. Linux systems are mainly employed in network computers and centralized processing, whilst Windows systems are used in user (console) and office environments.

6.2.2.2 Voice Communications System

An essential communications tool amongst experts participating in a mission is the voice system. Using terminals digital **voice conferencing system** (Figure 6.2.5), users inside and outside of the control center can be interconnected to subject- or task-oriented conferences. Participants in such conferences can be experts located in a control room onsite in external control centers or ground stations, or even astronauts in orbit. Terminals are configurable and may be connected to various conferences in different modes (listen only, talk enabled). The digital conferencing node allows for the recording of selected conferences and time-controlled selective playback.

6.2.2.3 Video System

Video has evolved as a standard communications tool in control centers. Besides supporting public relations activities, video is also utilized as an operations support tool (especially in the area of human space flight).

The technology trend is toward **digital systems**; video signals are distributed as digital streams via the data network and may be accessed throughout the control center and on request also in external partner facilities.

GSOC supports human space flight missions using digital video. Unmanned satellite operations are still supported via an analog video system installed around 1990. Special interfaces allow for cross-support of the two systems.

6.2.2.4 Power Supply

Besides the communication system connecting a control center to the outside world, energy supply is one of the most important support systems. A typical approach for a reliable power supply is described below.

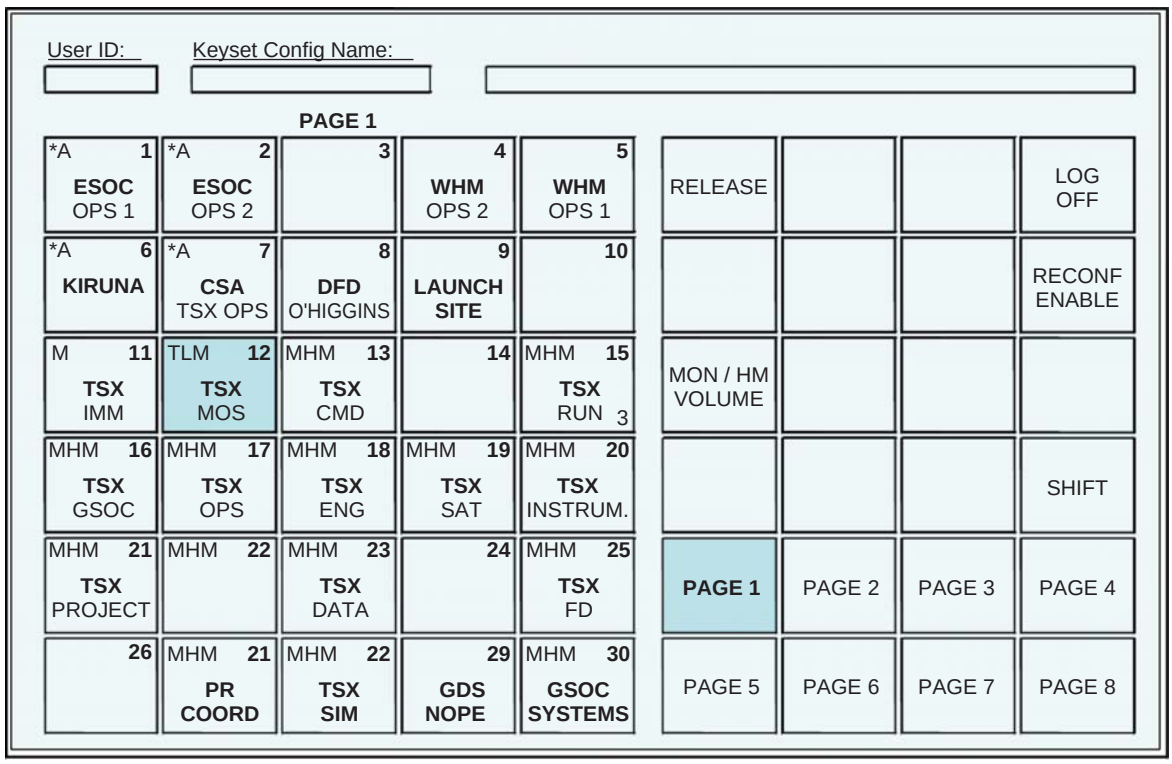


Figure 6.2.5: *User interface of a voice communications system (Source: DLR).*

All systems essential for operations are connected to an **uninterruptible power supply** (UPS). Users of electrical energy are fed from a so-called online supply. This supply draws energy from a set of batteries which are buffered from the normal power grid. This makes it possible to bridge interruptions of the grid lasting up to 10 minutes, as well as short disruptions. Within 30 seconds of the occurrence of an interruption, a **diesel-powered generator** takes over the function of the mains grid. This diesel system is able to run permanently which is important for long-duration grid failures, and limited only by the fuel supply.

The normal power supply (grid) is provided via two independent connections to the local energy company's high-voltage network (Figure 6.2.6).

All control center-internal subsystems are supplied with full redundancy.

6.2.2.5 Access Control and IT Security

Although control centers are usually hosted on the premises of existing security-controlled areas, additional local-level security measures are necessary to protect access to operations areas and systems.

Areas and rooms are secured via electronic badge-controlled door locks; critical areas are protected by additional fingerprint readers. During critical high-activity phases additional guard personnel are on duty. All critical areas and entry doors have video surveillance.

IT security is a complex matter which needs to be customized to mission requirements. **Firewall and proxy systems** are divided into various security tiers. Networks are also split into different security levels. The operations area is allocated to the highest security level and divided into projects and subprojects. Access to systems is password protected and implemented via the well-known **Lightweight Directory Access Protocol** (LDAP) which specifies and controls access rights.

Access to workstations is also controlled via card and fingerprint readers.

6.2.3 Software Components

Within a control center many specialized software systems are developed to support the operations

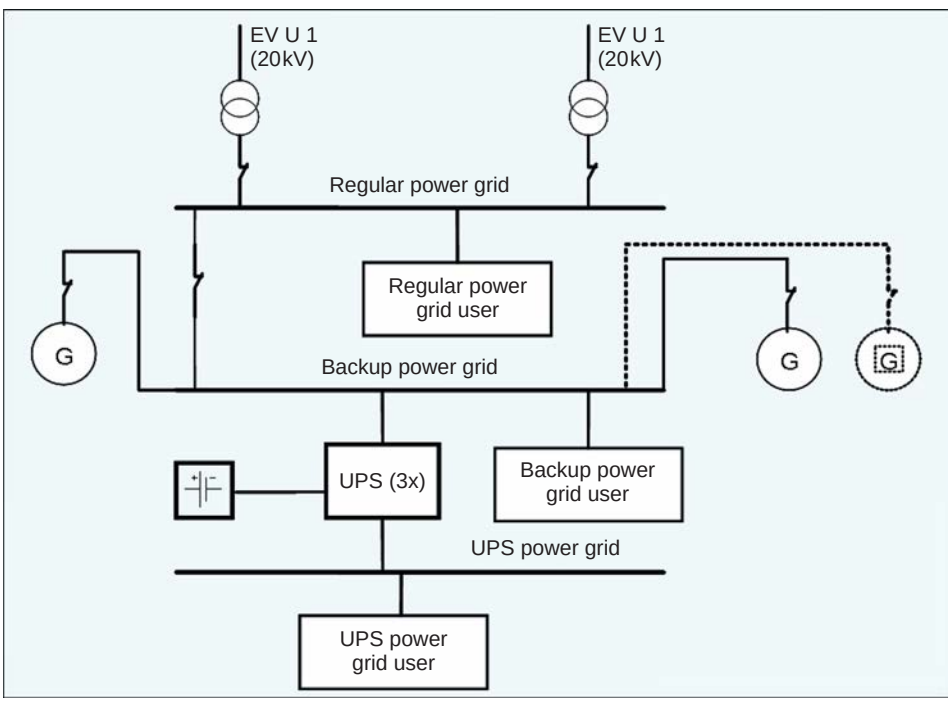


Figure 6.2.6: *Control center energy supply system (EVU = energy provider; G = diesel generator) (Source: DLR).*

teams. The most important ones are described in the following.

6.2.3.1 Telemetry and Command System

The **telemetry and command system** is the key element for satellite monitoring and control. In the early days separate software tools were implemented and used. Since 1995 integrated software systems have been made available, able to handle both tasks simultaneously. Examples are SCOS-2000, developed by ESA, the American EPOCH or the German MCS which was derived from the development phase of the ISS/Columbus project.

The following capabilities are common to all systems:

- **Processing** and archiving of telemetry data.
- Data distribution to users within or outside the control center.
- **Flagging** and alarming when single telemetry parameters violate limits, or generating summary alarms.
- Real-time as well as time-tagged commanding.

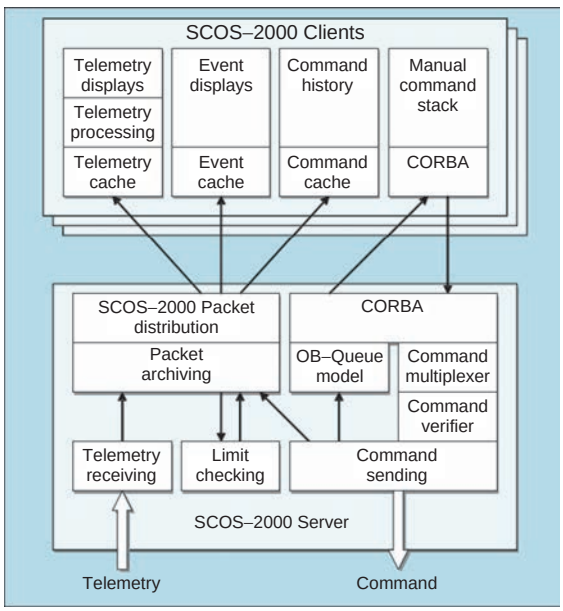


Figure 6.2.7: *SCOS-2000 system architecture (Source: DLR).*

- Logging of all transmitted commands.
- **Managing/displaying** the commands which are to be sent (Figure 6.2.8).
- **Autonomous sending** of commands relating to telemetry changes.

The rules specifying how commands are to be packed and how telemetry parameters are to be unpacked are defined in a special mission database. All allowed ranges for the different parameter values are also stored; these often depend on the mission phase or the overall status of a unit or subsystem. In order to achieve high reliability, two systems run in parallel (**hot backup**) and a switchover to the redundant system is accomplished in less than a minute. A third system is provided for critical mission phases (**cold backup**).

The system allows a **data recall (replay)** which is of special importance given the short contact times for Earth observation satellites, or the longer periods of no contact for probes. It is also required that data which cannot be received in real time and has to be dumped from the mass memory of the probe or the satellite has to be downloaded at a high data rate, processed and displayed at a higher than normal playback rate. Most of the telemetry and command systems like SCOS-2000 (Figure 6.2.8) support the relevant CCSDS (Consultative Committee for Space Data Systems) recommendations on **standardized telemetry and command packet formats** and ESA's packet utilization standards (PUS). The typical configuration of SCOS-2000 (Figure 6.2.7) is a distributed system based on a network of workstations. It may also be used in a minimized configuration as a single console solution (SCOS-2000 in-a-box) for simple missions or for test and training purposes.

6.2.3.2 Display System

Telemetry and telecommand systems do not always provide a user interface that can be handled in an easy and intuitive way. Therefore, and because of the use of different software tools for telemetry and telecommand data processing required by different projects, the need arises for one additional **user interface** layer. At GSOC the SATMON display system (Figure 6.2.9) is used, which provides many different types of data display in a very special intuitive way. Most common

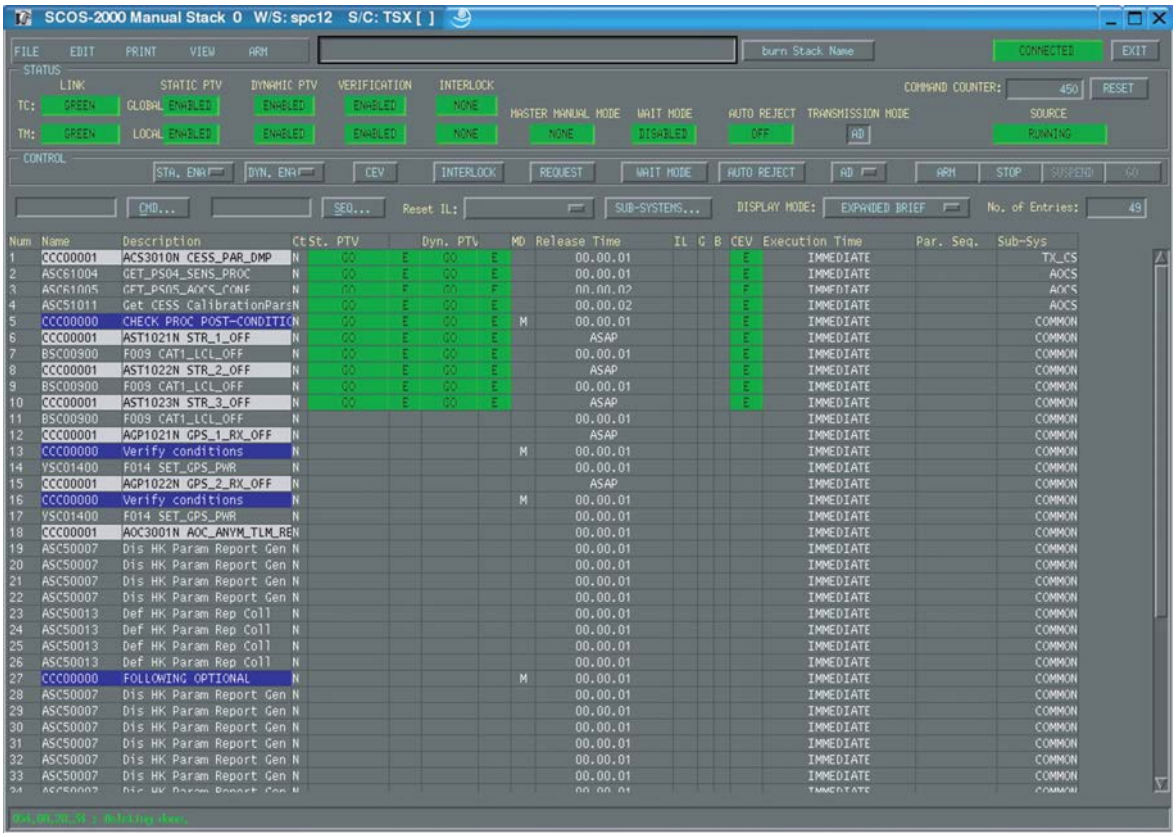


Figure 6.2.8: *SCOS-2000 command system (Source: DLR).*

are the standard alphanumeric displays of data in tabular form, and line plot functions with graphical data displays which allow for a better reaction to deviations by trend analysis. Additionally, a symbol plot representation of subsystems and equipment units can also be displayed in which the data values are more understandable in a graphical context together with other parameters.

Specific criteria of this software include the flexibility to be connected to different kinds of telemetry systems, the ability to integrate and display data from other sources, the provision of the telemetry display also via the Internet, graphical drag-and-drop configuration of the display pages layouts, hyperlinks between display pages, additional database information on all parameters through automatic pop-ups, derived parameters, limit checking, local data archiving, and replay. The display page content is adjusted to the window size. The display system runs on standard PC hardware under the Windows operating system, which is a cost-efficient solution for the high number of consoles in a control room.

Since 2005, thanks to the high performance of PCs, it has been possible to create **3D displays and animations**. At GSOC the software SatviZ (Figure 6.2.10) is used for selected projects; it is based on 3D models of the complete spacecraft or of single components. The specialty is that the current satellite configuration can be graphically presented on any console or on the big screen in the control room, reflecting the actual status and value of the telemetry. This provides additional visual support for the operations team, for example visualization of the satellite attitude.

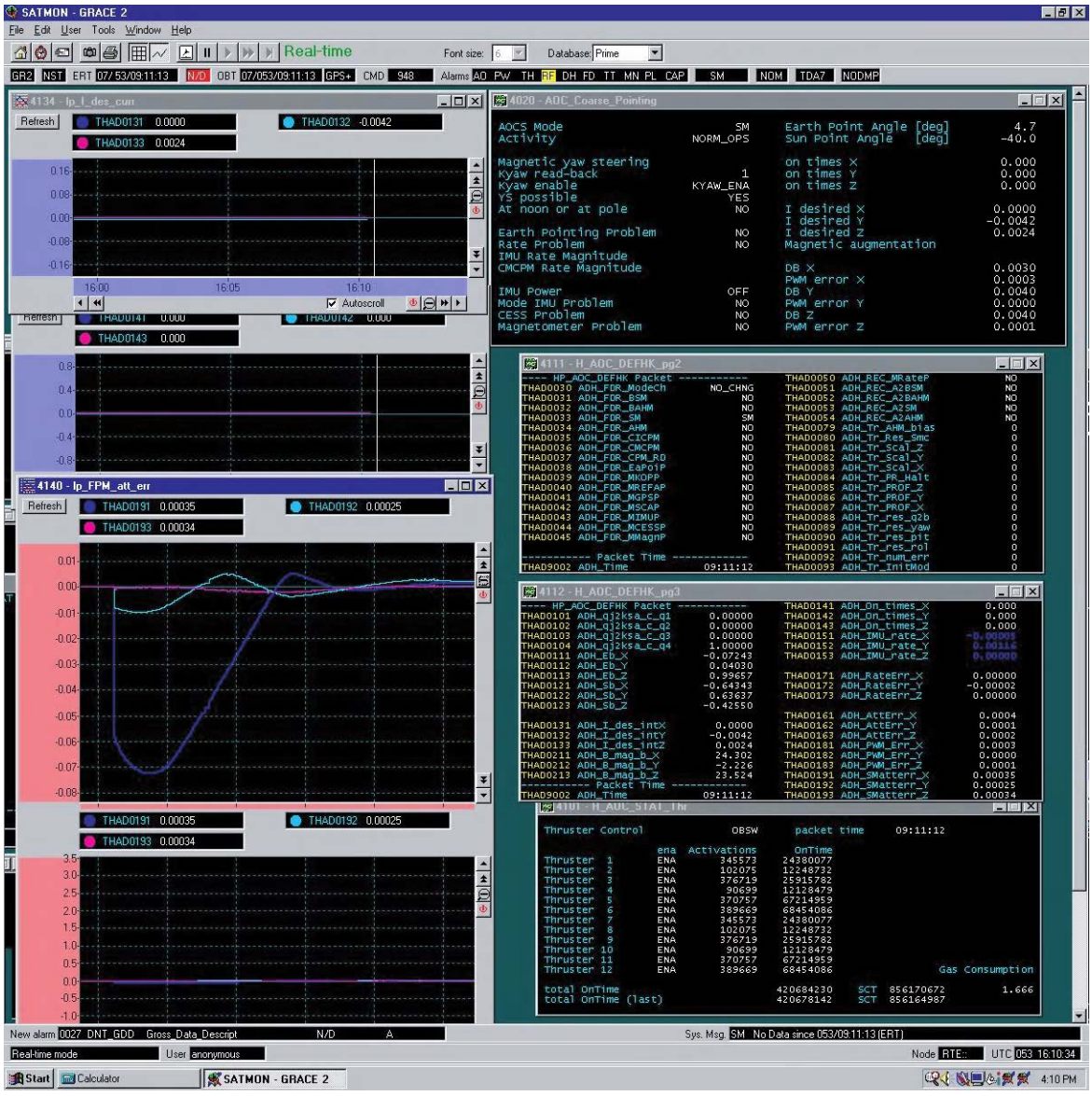


Figure 6.2.9: *Display system SATMON (Source: DLR).*

6.2.3.3 Mission Planning System

In order to prepare the detailed mission operations plan, **software tools** are deployed in many cases to help resolve the sometimes conflicting requirements without violating the given boundaries and constraints.

The software tool PLATO used at GSOC consists of four main components: a relational database, a graphical editor, an automatic planner and a Web interface.

The **database** is the core of the PLATO software. Here all information and parameter descriptions are stored. The generated plans are stored in the database as well. As the relational database MS Access has been chosen, since it allows standardized data exchange with other programs.

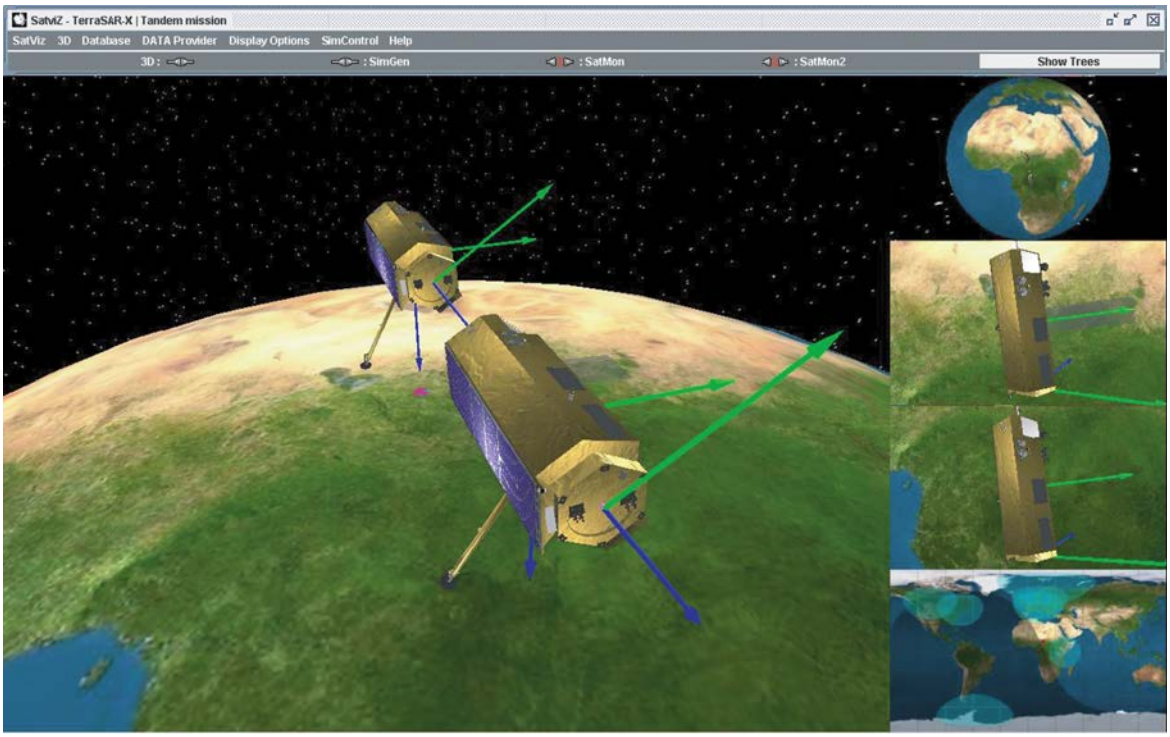


Figure 6.2.10: *The 3D display system SatviZ (Source: DLR).*

A **graphical editor** (Figure 6.2.11) provides several possibilities to define tasks in an intuitive way and to link and transfer them to other software systems. The graphical editor gives a basic overview as well as displaying detailed information. Some display formats can be created even during real-time operations.

The main task of the **automatic planner** is to generate a conflict-free plan. All interdependencies among the tasks and resource constraints are considered. The underlying algorithms originate from **constraint satisfaction problem theory** (CSP). This allows a solution to be found for complex problems within a reasonable processing time. Since the algorithms can be adapted to user requirements, PLATO can be used for different problem areas and applications. A server version allows several users to work in parallel. Tasks can be displayed hierarchically in three levels, the lowest level being the one which uses the resources. Different types of resources can be defined. PLATO provides the advantage of accepting not only hard constraints for the task planning like "begin-after-end" or "begin-X-days-after-start," but also soft or "fuzzy" constraints like "choose machine A for that task if possible and only if absolutely necessary machine B or C" (see, e.g., Figure 6.2.12)

The plan can be distributed over the Web as soon as the planner is convinced that the plan should be executed. Authorized users have access to the current timeline via the **Web interface**. The timeline is the result of the planning process. It contains precisely defined execution times for each activity and thus one of the executable and conflict-free solutions of the defined problem. In parallel the timeline presents additional options for using the resources showing all the potential availabilities of resources, constraints and time dependencies which led to the calculated solution. In many cases a data file can be provided containing all the telecommands (derived from the telecommand database) that are to be uplinked to the spacecraft.

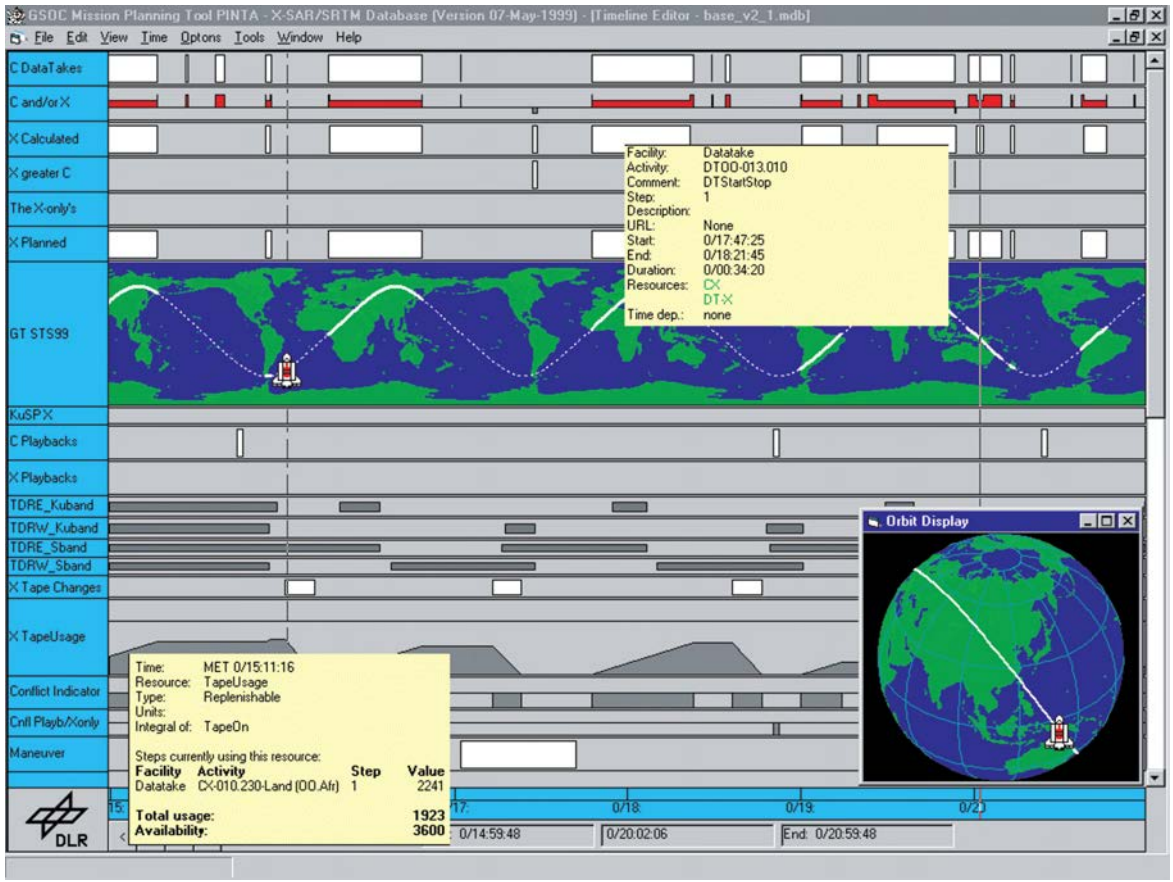


Figure 6.2.11: *Mission planning system (Source: DLR).*

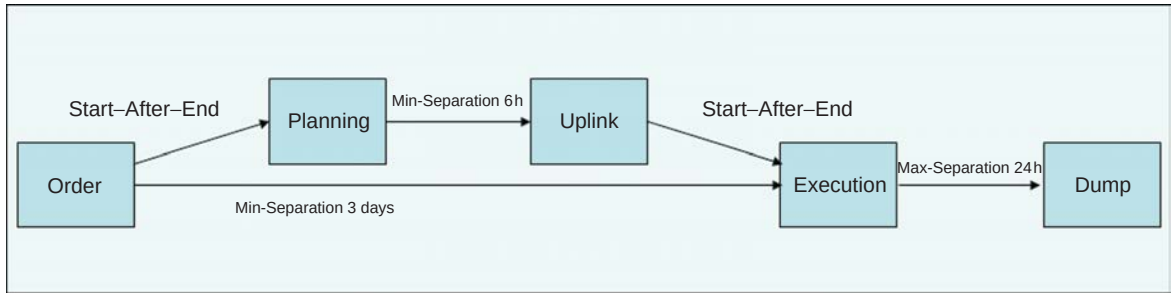


Figure 6.2.12: *Planning process (Source: DLR).*

6.2.3.4 Operations Support Tools

There is a lot of other software used for operations in the control center. First, the **time displays** should be mentioned. In space operations, UTC (Coordinated Universal Time, previously Greenwich Mean Time, GMT) is the basis for control centers all over the world. Nevertheless, it is good to know the local time of the partner(s) on the other side of the world. These times

are displayed on dedicated big screens in the control room or in small windows on the workstations, often together with specific-event time displays, for example the countdown or mission elapsed time (MET).

In many cases these event times are displayed together with the **current orbital position** (see Figure 6.2.13). From these orbit displays it can also be seen when a satellite is flying in sunlight or when the next communication contacts with a ground station can be expected acquisition of signal (AOS) or loss of signal (LOS).

Since 2002 another software area, the **OpsWeb**, has gained in importance in control centers. OpsWeb is an intranet which lists and provides under a single user interface access to the complete project documentation, all databases, and other supportive information like telephone lists and on-console plans.

The relevant information is gradually collected and validated by the operations team during the mission preparations phase.

6.2.4 Communications

Just like a spider, a control center "sits" at the hub of its communication lines. Without these connections

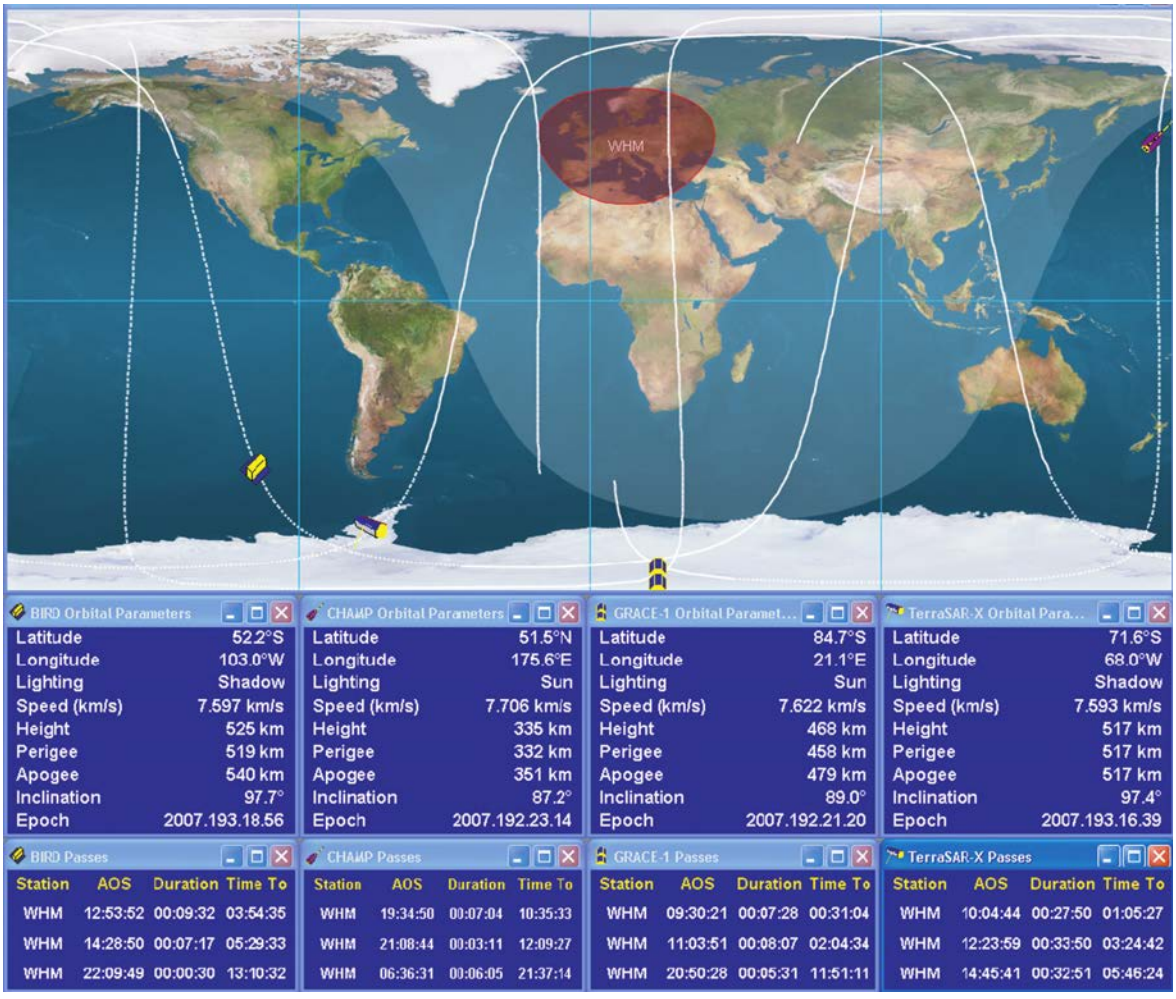


Figure 6.2.13: *Orbit display (Source: DLR).*

a control center is unable to function. In the early days dedicated links (pulse code modulated lines) had to be used. Now **digital links** and associated technologies are employed, tailored to the detailed **project requirements**.

These requirements define the type of connection (leased line, dialup line), which strongly depends also on the available communications budget of the project. The type of communications technology changes dynamically in this field. Currently, ISDN, ATM or IP over SDH are utilized.

6.2.4.1 Real-Time Data Transmission

Unlike the early days, real-time data transmission is not of the highest importance today. Previously real time was required because pulse code modulated data was not easy to store and had to be transmitted from source to destination upon receipt. In addition, computational intelligence on-board spacecraft was limited and control centers had to be included in the **control loops**.

Today, spacecraft have a high level of **autonomy**. Telemetry and telecommands are used to monitor and control spacecraft functions in an off-line fashion. Program and parameter loading as well as housekeeping data downloading are the main support functions.

6.2.4.2 Offline Data Transmission

Off-line data transmission is defined by **time-delayed** data transfer from control centers or ground stations (e.g., scientific data from experiments). If not prohibited by reasons of propriety and high volume, the public Internet or science intranets are used to support off-line data transmission.

6.2.4.3 Audio and Video Links

Audio transmission is in principle defined by the requirements of telephone systems. **Delays** of more then 150 milliseconds are disturbing in conversations and have to be avoided. This is, however, not feasible for long-distance interconnections due to physical reasons. Voice discipline and adherence to protocols help to avoid problems. Transmission of video and audio in parallel but using different transmission lines requires lip synchronization of the audio transmission. Video transmission is very susceptible to **variations in signal runtime**. Special buffer techniques help to avoid problems.

6.3 The Network of Ground Stations

Martin Häusler and Klaus Wiedemann

6.3.1 The Functions of a Ground Station

Operational communications between a control center (the spacecraft control and/or the payload control) and a spacecraft, in principle, are effected through a ground station. The transmission of user services between the ground and the satellite is accommodated on radiofrequency carriers.

6.3.1.1 Transmission from Ground to Satellite, the Uplink

For transmission on the uplink the ground station modulates the signals (*digital data*) received from the control center onto a **radiofrequency carrier**. The uplink transmission encompasses the transmission of telecommand signals from the ground station for the spacecraft platform, commands to the payload, signals for ranging of the space vehicle, and, in the case of communications satellites, the telecommunications services of the users. To this end, the signals are modulated onto the radiofrequency carrier; the carrier is then amplified and sent through a suitable antenna to the satellite.

6.3.1.2 Transmission from Satellite to Ground, the Downlink

The transmission from the satellite to the ground (downlink) is typically received with the same antenna that transmitted the uplink information. To avoid interference between transmit and receive signals, they are sufficiently separated in frequency on uplink and downlink.

The information transmitted on the downlink includes data on the condition of the satellite (**housekeeping data**), data on the configuration and performance

of the payload (**payload data**), signals for the **ranging** of the space vehicle, and, in the case of communications satellites, the **telecommunications** services.

The carriers received at the ground station are filtered and enhanced by **low-noise amplifiers** and then downconverted to an intermediate frequency, at which, after additional filtering, they can readily be **demodulated**. Today, the analog signals are generally converted to digital in analog/digital converters and then processed by special electronics (digital signal processors). Software processors can be used for data rates under 1 Mbit/s. After demodulation, the signals are restored to their original form. Due to the transmission from space to ground, the limited power on-board the satellites and the given size of the receiving antennas on the ground, the signals received are not free of errors. Extensive **processing** through filtering, synchronization and error correction is required before the data can be relayed to the control center via the appropriate communications channels. The basic communications architecture is different for the various types of applications.

The telemetry and telecommand data (TM/TC) is relayed between the ground station and the control center in real time. Navigation data, such as direction, distance and velocity of the spacecraft relayed from the ground station to the control center, and the orbit prediction data emitted from the control center to the ground station, is relayed off-line (Figure 6.3.1).

Remote sensing **Earth exploration satellites** (Figure 6.3.2) are programmed to dump the data they collect during the time with no ground station contacts to a preprogrammed receiving ground station. For this operation no uplink is necessary.

Communications satellites (Figure 6.3.3) have transponders which convert and amplify the carriers received from the ground and then retransmit them back to the service regions on Earth. These service regions can be large scale, up to global or very large areas (e.g., the USA), or smaller, geographically concentrated areas served by spot beams.

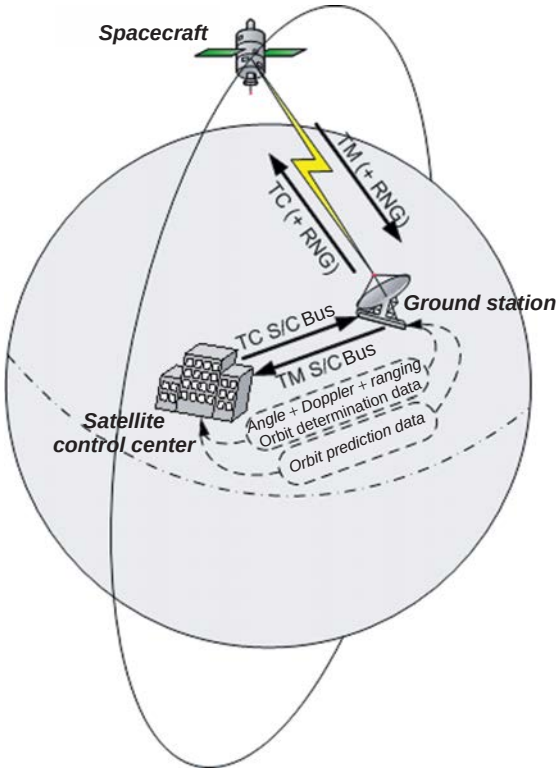


Figure 6.3.1: *The control of a spacecraft (Source: DLR).*

6.3.1.3 Telemetry

In space technology, telemetry is referred to as the transmission of data generated on-board the spacecraft, down to a ground station. This telemetry data can be of different types. The **payload data** of scientific experiments can be sensor data, images or radar data of different objects. The **housekeeping data** carries information about the status and health of the spacecraft such as the present charge of the batteries, the supply voltage, the temperature, the switch settings, how much fuel is left in the tanks or information on the attitude of the spacecraft. Often additional information is transmitted, for example about the orbital position of the spacecraft. This telemetry data is preprocessed at the ground station to be passed onto the control center. For safety reasons received telemetry data is stored and archived in the control center for several weeks, and always for the duration of the mission.

6.3.1.4 Telecommand

A spacecraft is commanded from the ground by **telecommands.** They allow the configuration of the

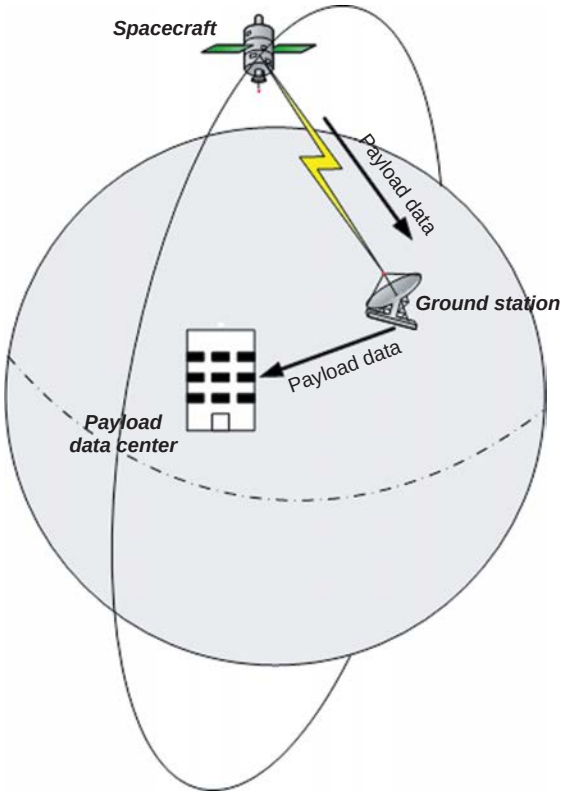


Figure 6.3.2: *Remote sensing (Source: DLR).*

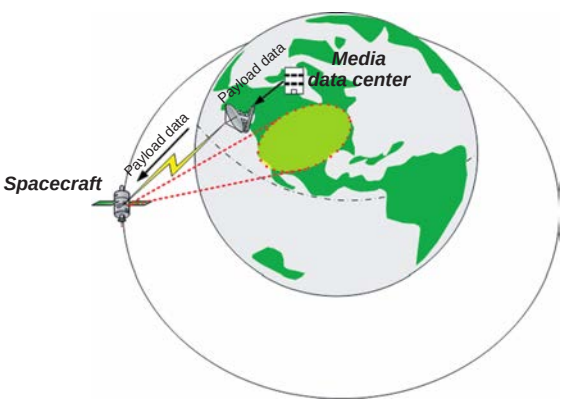


Figure 6.3.3: *Communications satellites (Source: DLR).*

spacecraft and its payload to be influenced. In addition, in accordance with the control station operational planning, the transmissions of telemetry and telecommand are preprogrammed. The telecommand to the spacecraft is used to program systems in the satellite to function autonomously, and also to perform complex orbital maneuvers using the satellite propulsion systems. State-of-the-art satellites and space probes today permit reprogramming or improvements to the on-board software, so frequent uploading of software has to be foreseen.

6.3.1.5 Tracking

In space technology, **tracking** refers to all activities concerned, in the widest sense, with the tracing of space vehicles, including procedures to assess and measure their orbits, using steerable antennas on the ground. Different techniques are employed to determine the exact orbits.

With angle measurements, angular data can be obtained relative to the location of the ground station. For low-orbiting satellites, this is performed with special antennas (autotracking antennas with monopulse or conical scan feeds). For the tracking of geostationary satellites, step track antennas can be used.

Range measurements relative to the location of the ground station can be made using a transponder in the satellite, if available, which loops uplink signals back to ground. Measuring the latency of the loopback signal, the distance between the ground station and spacecraft can be determined. This technique is used both for satellites and space probes.

The flight velocity of the satellite relative to the ground (the ground station) can be determined by measuring the Doppler frequency shift (range rate measurement) between the transmitted and received frequencies. For this procedure a transponder onboard the spacecraft is required also, for which, in this case, coherency between the received and transmitted frequency in phase and amplitude is required and must be known exactly. The measurement of the flight speed is used for remote sensing satellites and space probes and for geostationary satellites in transfer orbit.

For the determination of the spacecraft's position, navigation satellites are used, such as, presently, GPS. This works well when the orbit is suited to sufficient reception of such navigation signals. This is the case for satellites in LEO. Transmission of this data to Earth eliminates the need for orbit determination through the ground station.

6.3.2 Site Selection of a Ground Station

Appropriate site selection of a ground station is critical for the operation of the antennas free of radio interference. Several aspects need to be taken into consideration in selecting a site, such as the terrain; the required property size for placing the antennas; the ground structure and soil chemistry; the available infrastructure in terms of road access, electricity, water and particularly communication trunks; and the radio interference at the site. The importance of these aspects depends on the exact purpose of the ground station to be built. If it is to serve only geostationary satellites, there are different requirements than for stations serving satellites in LEO, which will remain in view of the ground station for only about 10 minutes during one pass.

In this context, it is important to determine the minimum **elevation angle** under which the satellite operation can be supported. Nearby hills and mountains help to block radio interference received from distant sources, but they limit the view angle of the satellite and thus shorten the time that a satellite is in the field of view of the ground station. To assess this situation, a **horizon profile** should be generated for the ground station; an example is depicted in Figure 6.3.4.

The figure illustrates how the surrounding terrain dictates the horizon. Between 120° and 300° azimuth a clear view of a satellite is given for elevations above 2° except for nearby obstacles. Below 2° satellite operation is not possible. For operation with LEO satellites, an unobstructed view of the entire sky is required. Therefore, obstacles should be in the far distance, and a large area is needed for the antenna. In the figure there are some obstacles shown which need to be taken into consideration in the operation of the antenna, such as, at 200° azimuth, there is an antenna mast (a straight line in the diagram) extending up to an elevation of some 14°. It will not interfere with the operation, but will certainly cause a small loss.

An example computation of this type of situation is shown for an antenna with a diameter of 11 m, that is an aperture of 95 m^2. In the near field, the power is distributed cylindrically from the antenna. The antenna mast (used for antenna calibrations) with a width of 0.7 m has a surface of 0.7 m × 11.0 m = 7.7 m^2, with which it blocks the radiation of the ground station and which leads to some 0.4 dB reduction of the field of view. In that situation, 91.9% of the radiated power still gets through; the link still works if an appropriate link margin has been budgeted.

For operation with communications satellites it may be sufficient to have an adequate view of a given

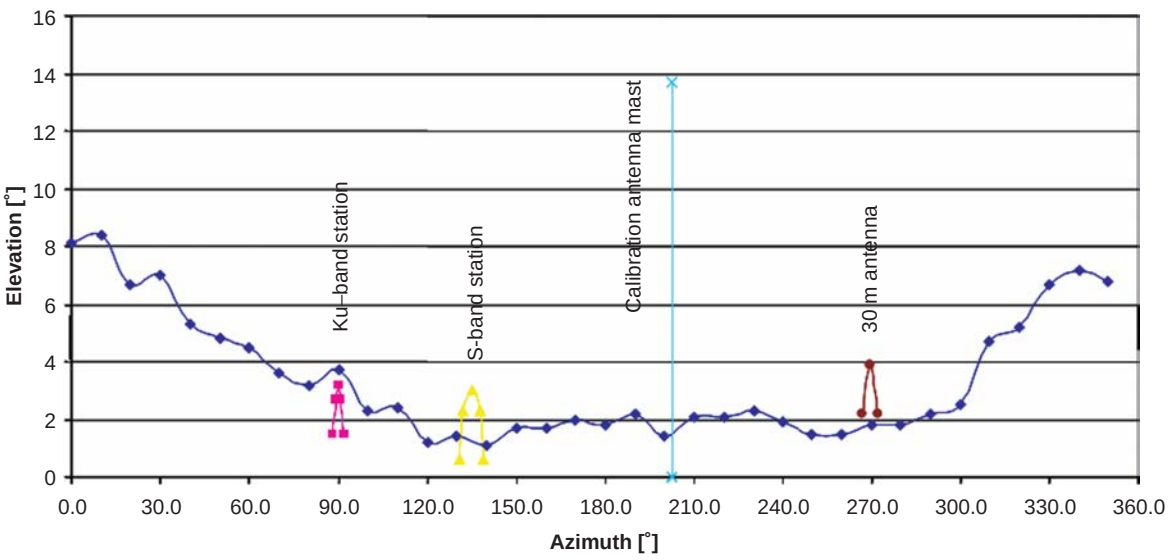


Figure 6.3.4: *Horizon profile of a ground station (Source: DLR).*

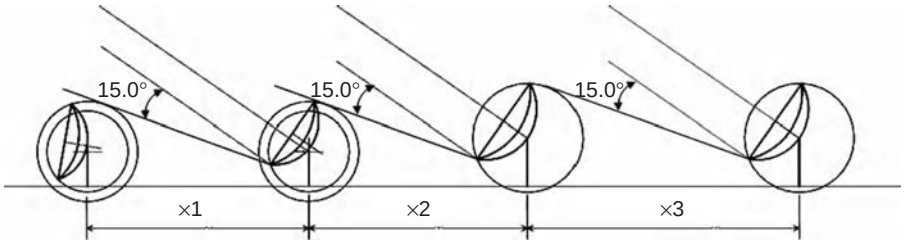


Figure 6.3.5: *Separation of antennas in a north/south direction (Source: DLR).*

geostationary satellite, or of the entire geostationary orbit as a maximum, so that for ground stations in the northern hemisphere obstacles of any type north of the station, such as buildings and mountains, will not cause problems. Often it is desirable to have several antennas for multiple simultaneous satellite links. When real estate is limited, the antennas must then be packed together in a small space. If the antennas can be arranged side by side in the east/west direction, they do not cause mutual interference. If this is not possible, they must be co-located in the north/south direction, sufficiently separated so they do not interfere with each other. Some antenna manufacturers recommend, for this case, to separate antennas for operation with the geostationary orbit so that an antenna has more than 15° angular clearance to the antenna in front of it (see Figure 6.3.5). One reason for keeping enough separation is the generation of passive intermodulation (PIM).

Another aspect to be taken into account when placing a ground station is the **soil condition**. Coastal regions often have quite sandy ground, which leads to expensive foundations; they are quite similar to loamy soils in this regard. Typically, the manufacturer of the antenna will quote the required ground compression factor. In any case, it is advisable to obtain an expert opinion on the soil condition to assure a solid and stable antenna foundation. This is all the more important if the antenna is to be used for satellite orbit determination. Here, difference settings for the foundations are the predominant problem.

To have a favorable interconnect **infrastructure** is of great advantage. Ingress and egress for personnel and transport are quite essential, as are connections to water, sewage, electricity and the telephone network, and to high-speed data services. These conditions are typically given when there is a town nearby, although settlements in the neighborhood can also be of disadvantage because they can affect satellite operations adversely by causing radio interference.

Other sources of radio interference such as mobile radio and similar telecommunications systems can lead to curtailment of the usable frequency spectrum of the ground station.

6.3.3 Subsystems of Ground Stations for Orbiting Satellites

All ground stations have similar architectures and employ similar subsystems to link up to the satellite. Figure 6.3.6 shows a block diagram of a satellite ground station.

The antenna transmits all signals using the uplink (U/L) chain, and receives all signals with its downlink (D/L) chain.

6.3.3.1 The Downlink

Figure 6.3.6 displays three receive chains each with one low-noise amplifier (LNA), which feed into a **three-channel downconverter** (3 CH D/C). The downconverter transposes the receive frequency into the intermediate frequency (IF) and inputs it through a downlink matrix into the baseband equipment (TTC baseband). The IF used is always the same, so that all antennas and all carriers into and out of these antennas can operate with the same standard **baseband equipment**. In the baseband, the signals are digitized into bit sequences which are relayed via the network (LAN) to a telemetry processor for evaluation.

6.3.3.2 The Uplink

In the other direction, the commands are sent to the satellite, starting with the station computer sending

them via the network (LAN) to the baseband equipment, which, again, transposes them to the intermediate frequency. Then they are modulated onto a radiofrequency carrier which is passed on to the antenna where they are finally **upconverted** (U/C), high-power amplified (HPA) and inputted into the antenna for transmission.

The U/L and D/L matrices in the transmit and receive chains link the antennas to the baseband equipment. This allows flexible connection of different baseband equipment to an antenna, a feature of considerable utility in large ground stations.

6.3.3.3 Antenna Tracking

There are three types of antenna tracking techniques.

In the **monopulse tracking system** four feeds are arranged in the quadrants of a circle around the beam axis, which receive four different signals. The difference between the two upper feeds and the two lower feeds refers to the elevation delta, and the difference between the two feeds on the left and the two feeds on the right refers to the azimuth delta. These deltas are converted to offset voltages which are continuously available and can immediately be used in the antenna control unit (ACU) to automatically track the antenna on the target. The system is also referred to as **simultaneous lobing**.

The **conical scan feed** is moved mechanically and describes a cone. This generates an amplitude-modulated signal. This sequential signal can be evaluated and then used to keep the antenna pointed correctly.

The **step track method** is generally employed in fixed antenna systems working with GEO satellites. The antenna is continuously moved slightly left and right (azimuth) and then, similarly, up and down (elevation); the received signal amplitudes are compared to determine the maximal signal power and thus to optimally point at the satellite.

This conical scan and the step track technique are referred to as **sequential lobing**.

6.3.3.4 Low Noise Amplifiers

The first LNA behind the antenna output is very important. It must enhance the extremely low-power

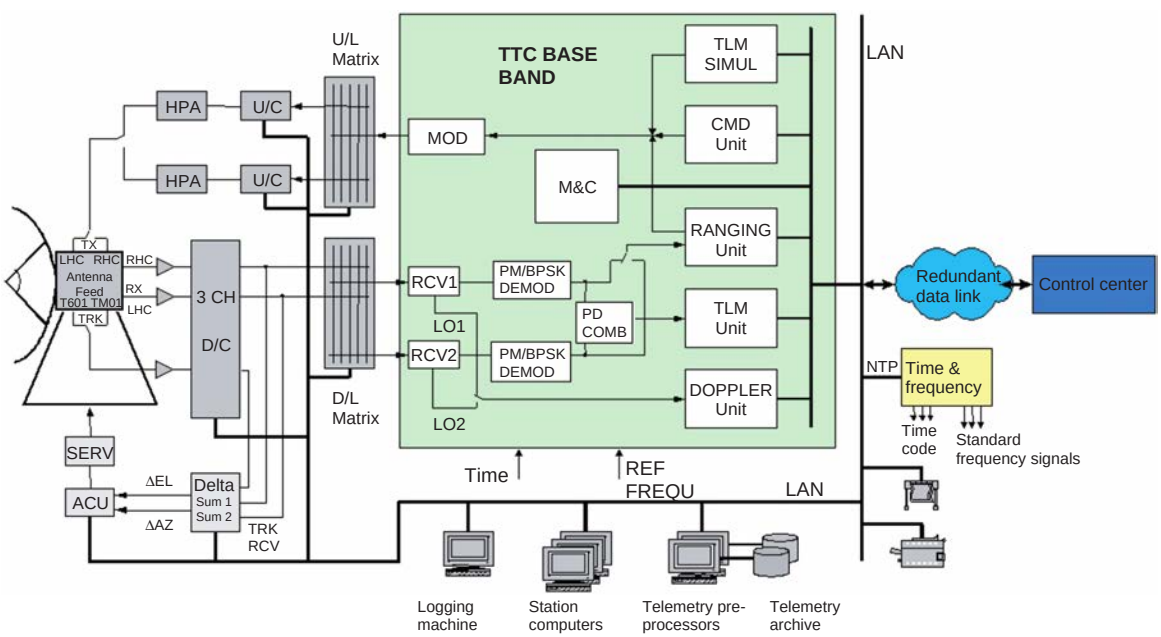


Figure 6.3.6: *Block diagram of a ground station with monopulse tracking for LEO satellites (Source: DLR).*

receive signal and, at the same time, not add noise to it. Thus, the relevant characteristic of the LNA is its noise figure, which defines the ratio of signal to noise at the amplifier input to signal to noise at its output. The LNA should be as close to the antenna as possible, without long cable connections or other components since the losses between antenna and LNA degrade the signal-to-noise ratio.

6.3.3.5 High-Power Amplifiers

To implement the required link budget for the uplink from ground to satellite, generally more than one **high-power amplifier** (HPA) is used, which enhances the output of the upconverter sufficiently to command the satellite. The gain in amplification is related to the antenna gain, which in turn is determined through the diameter and type of the antenna. The result is the **equivalent isotropic radiated power** (EIRP) of the ground station.

There are different types of HPAs: traveling wave tubes (TWTs), klystron amplifiers and solid-state power amplifiers (SSPA). These are briefly described and their advantages and disadvantages discussed.

The **TWT amplifier** is the most common power amplifier in space applications. Almost all satellites use this type on the downlink from satellite to ground. The tube requires a voltage for the heating of the cathode and a very high voltage of several kilovolts to accelerate the electron beam. The high degree of reliability, as required in satellites, results from the relatively simple peripheral equipment. TWT amplifiers also exhibit the high efficiency required in satellites which have limited power on-board, and they have a remarkably broad bandwidth (about one octave) without requiring special tuning to carrier frequencies. For this reason their use is also widespread in ground stations to amplify the uplink signals, since they require neither frequency tuning nor mechanical maintenance. One disadvantage is the decreasing power of the aging cathode. The output power of TWTs ranges from several watts to several kilowatts. Even at very high operating frequencies, the K-band (18–26.5 GHz), TWTs produce substantial power levels.

The **klystron** works similarly to the TWT, that is with a vacuum-filled tube, a cathode, an anode and a collector. The difference is that in klystrons the frequency is set by tuning cavities. This causes the electron beam which is generated in the klystron to be mechanically directed through the changes of the tuning sections. By narrowing or widening the chambers, the electron beam is tuned to the center frequency, the resonance. Klystrons use four to five chambers to generate a homogeneous transfer function. The bandwidth of the tuning is a function of the gain and the desired output power. For large bandwidths the gain factor is reduced, and vice versa. The main advantage of this HPA lies in the high radiofrequency power that it produces.

For that reason, klystrons are used in communications requiring high power, as well as in pulsed radars. It is reasonable to use them when the required output powers exceed the capabilities of TWTs, that is some kilowatts to several hundred kilowatts, whereby klystrons operate both at low frequencies as well as at the high end of the frequency spectrum, but they do require substantial effort for general servicing and tuning. Also, the higher output power leads to lower efficiency compared to TWTs, and it also requires complex cooling systems (e.g., water cooling).

A state-of-the-art type of amplifier is the **solid-state power amplifier** (SSPA). The continuous development of semiconductor technology has led to transistor amplifiers with acceptable efficiencies in the gigahertz range, though the power consumption increases exponentially with frequency. The output power at S-band (2–4 GHz) is presently some 100 W with GaAs semiconductors, with an efficiency of 30–40%. Power can be increased by paralleling amplifiers which then, of course, need to be tuned in phase. With power combiners, amplifiers can be paralleled also. These show a relatively large input attenuation, as a function of frequency and power, up to 1 dB and more. In summary, for higher frequencies and/or higher power, TWTs and klystrons are clearly superior to SSPAs. The only advantage of SSPAs is the short warm-up times – a tube amplifier requires several minutes for the cathode to heat up.

The **mounting of the HPA onto the feed system** of the antenna must produce minimum losses. Arcing in the waveguides caused, for example, by condensates can be avoided with nitrogen overpressure or dry air. The cost and mounting effort of waveguides are absolutely required for high power

levels at high frequencies. Only medium power levels can be transported by coaxial cables. Their flexible installation and low cost compensate the higher attenuation in this case.

6.3.3.6 Upconverter and Downconverter

Upconverters and downconverters transpose the frequencies and thus connect the baseband with the HPA or LNA on the receive side.

In the **upconverter** the modulated signal is transposed from intermediate frequency up to transmit frequency. The upconverter consists of a frequency transformation, a synthesizer and a power amplifier. The frequency translation typically is performed in two steps (dual conversion), to avoid intermodulation. The mixer frequencies are generated in digital synthesizers and synchronized with temperature-stabilized quartz oscillators or external station reference frequencies. The transmit signal then is bandpass filtered and enhanced in a driver amplifier to the input level required by the following HPA.

In the **downconverter** the received signal output of the LNA is transposed to an intermediate frequency suitable for the baseband equipment. Depending on the feed system used, single-channel or multichannel converters are employed. A multichannel converter is used in receive systems with a combiner and/or an autotracking system. The elements of the downconverter are similar to those of the upconverter, namely a frequency converter and a synthesizer.

The received signal is first bandpass filtered, then amplified and fed into the frequency converter. Again, the frequency conversion is done in two steps (dual conversion) to minimize intermodulation. The signal at the intermediate frequency is again bandpass filtered and is then inputted into the baseband equipment.

6.3.3.7 Baseband Equipment

The baseband equipment (e.g., CORTEX) can be considered as the heart of the ground station. It includes the telecommand system, the uplink modulator, the receiver, the combiner, the telemetry system, the Doppler measurement system, the ranging system, the simulator and data storage.

State-of-the-art baseband equipment consists of a powerful **industrial PC** equipped with special input and output cards for **uplink and downlink**. These uplink and downlink cards interface with the analog signals at intermediate frequencies (e.g., 70 MHz) and the digital PC. The downconverter output is transposed in a high-resolution analog/digital converter for further processing. In the other direction; the digital signal is digital/analog converted, sent through the upconverter and then to the uplink. To precisely format data in time, the baseband equipment has a **timer code card** which is slaved to an external station time. All frequencies generated at baseband level can be synchronized with the external station reference frequency. A network gateway is foreseen for monitoring and control as well as data transmission (e.g., Ethernet).

The **telecommand system** transforms the information sent from the control center or generated locally into a predefined protocol. There is a decision whether the command is to be handled in the start–stop (i.e., the modulation is turned off after each command) or in idle mode (an idle sequence is sent between commands to lock the bit synchronizer in the satellite).

The **uplink modulator** generates the intermediate frequency with the appropriate power level needed in the upconverter for the modulation used. Typical modulations on the uplink are frequency modulation (FM), phase modulation (PM) and binary phase shift keying (BPSK; see also Section 4.7.6). In addition, the modulation index for PM and the frequency deviation for FM can be set. For the input to the modulator, several sources are available: internal, the telecommand system (TC), the ranging measurement (RNG) and the simulator (SIM). For "nonstandard modulations" there is an external modulation port (AUX). A further option is the sweep generator in the uplink modulator to synchronize the uplink signal with the phase-locked loop (PLL) receiver in the satellite; the sweep range and the sweep rate can be set.

The intermediate-frequency signal from the downconverter is fed into the receiver. The receive frequency is set and the sweep range defined. A **loop filter bandwidth** is set corresponding to the expected signal-to-noise ratio. The receiver comes in two modes, a video mode and a PCM mode. The **video mode** is used for

frequency- or phase-modulated signals. A typical application is the transmission of housekeeping data of low data rate and bandwidth. For the transmission of high data rates with large bandwidths such as for sensor signals (optical sensors, magnetometers, radar sensors, etc.), generally the **PCM mode** is used with BPSK, QPSK, OQPSK or PCM/PM modulation.

The satellite downlink signal is polarized linearly (vertical or horizontal) or circularly (right hand or left hand). Due to impurities the signal will have a certain component in the orthogonal polarization. A **combiner** is used to receive a signal with the maximum signal-to-noise ratio in the ground station which receives both polarizations, downconverts them to the intermediate frequency and inputs them to two receivers. The combiner then passes the better signal to the telemetry system or combines the two signals. The use of this combiner can improve the signal-to-noise ratio by up to 2.5 dB.

In the **telemetry system**, arranged after the receiver or combiner, the signals are PM, FM, BPSK, QPSK, OQPSK or PCM/PM **demodulated** and then the bit sequence with the PCM coding (BP-L/M/S or NRZ-L/M/S) is checked and synchronized with the nominal bit rate. Coded signals then go through the decoder. Common coding techniques are Viterbi, Reed–Solomon, scrambling and combinations of these. The bit-synchronized and decoded signal is now checked for the sync word and the frame length. A CRC check can also be performed. The processed data frames, equipped with a very precise time stamp, can be stored internally for future processing, or transmitted to the control center, in real time.

An integral part of the receiver is the **Doppler measurement system** which continuously measures the receive frequency and determines the Doppler rate. The Doppler rate is a measure of the satellite's velocity and contributes to orbit determination.

The **ranging measurement system** includes a ranging modulator and demodulator, serving different ranging standards such as the tone ranging standards of ESA, ESA-like or INMARSAT, as well as ESA code ranging. For the desired standard, the major and minor tones, the amplitudes of the tones and the code length for code ranging are set. To avoid ambiguities in the range measurement, the minimum and maximum distances to the satellite are used as input.

Before the ranging measurement proper, the latency of the station equipment is calibrated. The ranging signal generated in the baseband equipment is modulated, the intermediate-frequency signal upconverted to the transmit frequency, enhanced in the HPA to the required power level and sent by the antenna, which, for this purpose, does not aim at a satellite but at zenith, at 90° elevation. A mixer in the antenna, the "reflective converter," transposes the transmit signal to the downlink frequency. The signal is then led through the receive chain, starting with the LNA. The downconverter mixes it with the intermediate frequency and it is synchronized and then demodulated in the ranging demodulator. The ranging measurement system now determines the latency of the signal. The station latency is taken into consideration in the ranging of the satellite.

The signal flow in the **ranging measurement** proper is the same as in the calibration except that the reflective converter is not switched on and the antenna is pointed at the satellite. The satellite must be set to the ranging mode; that is, the uplink ranging signal is added to the telemetry data for the downlink and sent off. In the ranging system, again, the latency of the signal is measured. Subtracting the station latency (as measured in the calibration) and the satellite latency (as supplied by the satellite manufacturer) from the total latency, and then dividing the result by two, leads to the latency between ground station and satellite.

A simulator is used to check the receive system and telemetry system. The uplink modulator output is connected to the receiver via an attenuator. The simulator is configured so that frame length, sync word, coding, bit rate, modulation, modulation index and receive frequency correspond to the settings in the receive system and in the telemetry system. The data in the frame can be generated in real time or can be pseudo-random data generated by the simulator.

6.3.3.8 Frequency and Time Reference

The American GPS (Global Positioning System) is the basis for deriving both the time and the reference frequency (Figure 6.3.7). In satellite operations, it is commonly used as a **time and frequency standard**. The exact time is required for the reception of telemetry data, the transmission of commands (time-tagged commands),

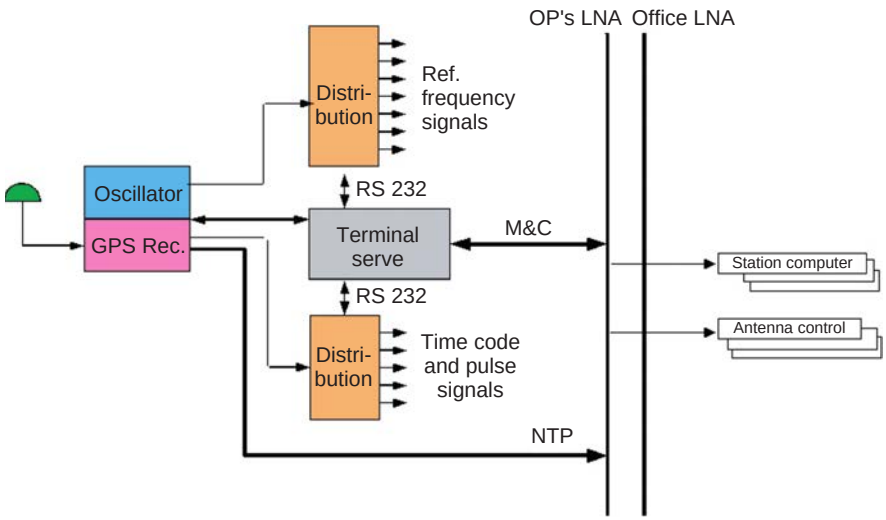


Figure 6.3.7: *Time and frequency system using GPS (Source: DLR).*

the tracking of the antenna and for orbit determination. Baseband equipment and up- and downconverter are supplied with reference frequencies.

A GPS receiver processes the received **time signal** and generates the time codes (NASA36, IRIG-B, 1PP), which can then be distributed within the ground station. In addition there is an oscillator integrated in the GPS receiver which is the basis for a **reference frequency**. Commonly, different reference frequencies are generated (at 1, 5 and 10 MHz) and distributed.

6.3.3.9 Data Network

The data network of a ground station should consist of multiple, physically separated **local area networks** (LANs). With a structured cabling in the different buildings, each of these LANs and patch panels can access any port in the various rooms of the building. The connection between the building and the antennas is separate for each LAN, for example over a multimode 50/125 µm fiber optic link, depending on the distance between the buildings (gigabit links). It is reasonable to physically separate the several network areas within a ground station for security reasons, which can be quite important in some applications. Basically, the following three areas can be defined.

The operational system (**OPS-LAN**) interconnects all buildings and subsystems in the ground station, including the antenna steering, transmitter and receiver, up- and downconverter, etc. Station computers control the different antenna systems and their subsystems. The OPS-LAN is connected to the control center via routers and firewalls using redundant, dedicated data lines. Telemetry, command and orbit data is exchanged over this link. The transmission protocol is TCP/IP. An NTP (Network Time Protocol) server in the OPS-LAN provides a GPS-synchronized UTP time reference for all clients. At present, all systems in the OPS are linked in a virtual LAN (VLAN). Additional VLANs for future extensions are readily possible.

The computers of all personnel in the ground station are networked in the **office communication system** (OCS). The file and print servers provide data exchange in defined working groups. The OCS of the ground station is connected to the OCS of the control center via routers and firewalls using a separate, dedicated data line to allow personnel at the ground station to access the file server of the control center. Access to the Internet and to mail servers is also handled over this link. Virus scanners and firewalls are employed for safe transmission.

The **external system** (EXS) allows access for external customers to subsystems of the ground station.

The external control center can temporarily operate subsystems of the antenna facility such as the CORTEX system, for example, in direct responsibility, via leased lines or via ISDN. Such subsystems are separated from the OPS and integrated into the EXS during entire passes, as agreed with the external customer.

All components of the three data networks (switches, routers, firewalls, etc.) can be configured via a virtual management network. Every single connection should be kept under surveillance to allow rapid detection and correction of failures in the network.

6.3.4 Link Design Aspects

The design and dimensioning of a transmission is customized to the individual service at hand. It is quite different for fixed and mobile applications and particularly for different transmission rates and service availability requirements. The **physical availability** of the link depends on the frequency range and the climate zone of the ground station, both on the uplink and on the downlink side.

The **link** – the radio connection between two points relayed via one or more satellites – is the total link from the transmitting ground station to the satellite (uplink), possibly on to another satellite in orbit (via intersatellite links) and back to the receiving ground station (downlink). It does not include the terrestrial interconnects from the ground stations to the users. The handover point is in the ground station, at the channel encoder input, and, at the other end, at the channel decoder output.

For successful transmission over this extreme radio link the transmit power and the noise power must be dimensioned on each partial link such that the desired signal quality and availability of service can be guaranteed for the resulting total transmission.

The **signal availability** p as a percentage of time is defined for the worst month in the year; the percentage values correspond to the following outage times:

$p = 98.0\%$	$\rightarrow$	14.6 hours/month
$p = 99.0\%$	$\rightarrow$	7.3 hours/month
$p = 99.9\%$	$\rightarrow$	43.8 minutes/month
$p = 99.99\%$	$\rightarrow$	4.38 minutes/month
$p = 99.999\%$	$\rightarrow$	26.3 seconds/month.

Planned downtimes of the ground stations for maintenance etc. are organized such that they do not lead to nonavailability of service.

The signal quality of the transmission of digital signals is measured by the **bit error rate** (BER) or the **packet error rate** (PER), whereby

- a given $BER = k \cdot 10^{-m}$ means that k bits of 10^m bits are in error;
- a given $PER = m \cdot 10^{-m}$ means that there are $m \cdot 10^{-m}$ errors in a packet of size n bits (n is the number of bits per packet).

Typically $m = 3$ for uncompanded voice, $m = 5$ for general data and $m = 7$ or higher for critical data.

The relationship between BER and PER is as follows:

$$\begin{aligned} PER &= 1-(1-BER)^n \\ &= 1-(1-n\ BER \pm n(n-1)\ BER^2/2 \\ &\quad \pm \ldots \pm BER^n) \\ &\approx n\ BER \end{aligned} \tag{6.3.1}$$

For common packet sizes and BERs, the PER is equal to the BER multiplied by the number of bits in the packet.

For the **design of the transmission link**, a suitable frequency band has to be determined as well as an antenna size for the ground station such that the power radiated on the uplink in conjunction with the figure of merit on the downlink leads to the required link quality.

The relevant parameters and quantities of the link are given in the following sections.

6.3.4.1 Frequency Ranges

The **frequency spectrum** from 7 kHz to 1000 GHz is divided into ranges designated by industry (Institute of Electrical and Electronics Engineers (IEEE), USA) and the military (German armed forces, NATO, etc.) as in Table 6.3.1.

None of these bands covers the entire frequency range up to 1000 GHz. The military sequence of stages seems logical but is impractical since several important frequency ranges such as Ku-band (10–19 GHz) and Ka-band (20–30 GHz) fall into the same SHF

Industry (IEEE)		Military (Germany)		Military (NATO)	
I-band	0–0.2 GHz	ELF	30 Hz – 3 kHz	A-band	0–0.25 GHz
G-band	0.2–0.25 GHz	VLF	3–30 kHz	B-band	0.25–0.5 GHz
P-band	0.25–0.5 GHz	LF	30–300 kHz	C-band	0.5–1.0 GHz
L-band	0.5–2 GHz	MF	0.3–3 MHz	D-band	1.0–2.0 GHz
S-band	2–3 GHz	HF	3.0–30 MHz	E-band	2.0–3.0 GHz
C-band	3–7 GHz	VHF	30–300 MHz	F-band	3.0–4.0 GHz
X-band	7–10 GHz	UHF	0.3–3 GHz	G-band	4.0–6.0 GHz
Ku-band	10–19 GHz			H-band	6.0–8.0 GHz
Ka-band	19–36 GHz	SHF	3.0–30 GHz	I-band	8.0–10 GHz
Q-band	36–46 GHz			J-band	10–20 GHz
V-band	46–56 GHz			K-band	20–40 GHz
		EHF	30–300 GHz	L-band	40–60 GHz
W-band	56–111 GHz			M-band	60–100 GHz

Table 6.3.1: *Frequency band designation.*

range (3–30 GHz) and thus cannot be differentiated (this problem is often solved by calling the 20–30 GHz band, incorrectly, the EHF range).

Caution is also needed with regard to the abbreviation HF, which in German stands for radiofrequency (3–30 GHz) and in English for short-wave frequency (3–30 MHz). The NATO designations are particularly unfortunate since they used the same letters as industry (IEEE), leading to misinterpretations at L-band and C-band (only the K-band is approximately the same).

The **choice** of a suitable **frequency band** for a given application depends on the type of service required and the availability of satellite capacity, which can be leased. For communications, space operations, broadcast (TV), remote sensing and reconnaissance, research and amateur radio some 20 **radio services** have been defined by the ITU[1] (Table 6.3.2).

The **frequency allocations** are defined in the three global ITU regions depicted in Figure 6.3.8: Europe, Africa and GUS (Region 1), the Americas (Region 2) and the rest of the world (Region 3).

The ITU endeavors to harmonize allocations globally, but often sufficient consensus can only be achieved regionally (which is often quite acceptable, actually). So to partition the world into these three regions is a pragmatic approach to the universal regulation: what is of advantage in Asia may or may not be advantageous in Europe.

The orbital positions and the radio services are handled by the ITU-R[2] on the basis of the Radio Regulations (RR), which, according to Article 4 of the Convention, are legally binding for the member states.

The ITU-R holds a World Radiocommunication Conference (WRC) every three years, which is attended by all member states, who then agree on the use of the frequency spectrum and on procedures and measures to administer it. The apportioning and use of the spectrum can be:

- an allocation of a frequency band to a service, by the ITU
- an allotment of spectrum to developing countries, by the ITU
- an assignment of a frequency band to an operator, by the national administration.

The assignment of a frequency band to an operator is handled by the national administration,[3] which, however, if the reach of the radio signals exceeds the

[1] International Telecommunications Union (Geneva), the UN organization globally administrating the frequency spectrum and the orbits.

[2] ITU-Radiocommunications sector.

[3] In Germany the Federal Network Agency (BNetzA im BMWA), Mainz/Berlin/Bonn.

Table 6.3.2: *Satellite radio services per the ITU.*

RR	Service type	Abb.
1.21	Fixed satellite service incl. feeder links for other space communication services	FSS
1.22	Intersatellite service	ISL
1.23	Space operation service, spacecraft tracking, telemetry and telecommand	SOS
1.25	Mobile satellite service, feeder links	MSS
1.27	Land mobile satellite service	L-MSS
1.29	Maritime mobile satellite service with ships; survival craft and EPIRBs	M-MSS
1.35	Aeronautical mobile satellite communications with aircraft, survival craft, EPIRBs	A-MSS
1.36	Aeronautical mobile satellite (route), safety and regularity of flights on civil air routes	A-MSS (R)
1.37	Aeronautical mobile satellite (off-route), communications for flight coordination off civil air routes	A-MSS (OR)
1.39	Broadcasting satellite service, for direct reception	BSS
1.41	Radiodetermination satellite service, incl. feeder links	
1.43	Radio navigation satellite service, incl. feeder links	RNSS
1.45	Maritime radionavigation satellite service; stations on ships	
1.47	Aeronautical radionavigation satellite service, stations on aircraft	
1.49	Radiolocation satellite service used for radiolocation, incl. feeder links	
1.51	Earth exploration satellite service, between ground stations and space incl. feeder links	EESS
1.52	Meteorological satellite service for Earth exploration	
1.54	Standard frequency and time signal satellite service incl. feeder links	SFS
1.55	Space research service for scientific or technological research purposes	SRS
1.57	Amateur satellite service radiocommunications using Earth satellites	ASS

national borders, hands it on to the ITU for international coordination.

The European Commission (EC) is building up a European frequency administration, with the burden that only 25 of the some 50 states in Europe are members of the EU.

6.3.4.2 The Size of the Antenna

The size of the antenna is essential for transmission and reception, although not solely. On the transmit side, the determining factor is the **product of transmit power and antenna gain, the effective-to-isotropic radiated power** (EIRP), and on the receive side it is the ratio of antenna gain g to the receive system noise temperature t (g/t or, transformed into decibels, the G–T), referred to as the figure of merit of the receive system.

When parameters are given in capital letters it means that they come in decibel values, whereby

$$X = 10 \cdot \log(x) \quad \text{in dB} \tag{6.3.2}$$

The transformation into decibels makes sense since parameters such as free space attenuation and, particularly, the Boltzmann constant k needed to determine the noise power are quite difficult to handle in the physical domain:

$$k = 1.38 \cdot 10^{-23} \text{ J/K.}$$

But, with $K = 10 \log(k)$, we get

$$K = -228.6 \text{ dB J/K}$$

which can readily be employed.

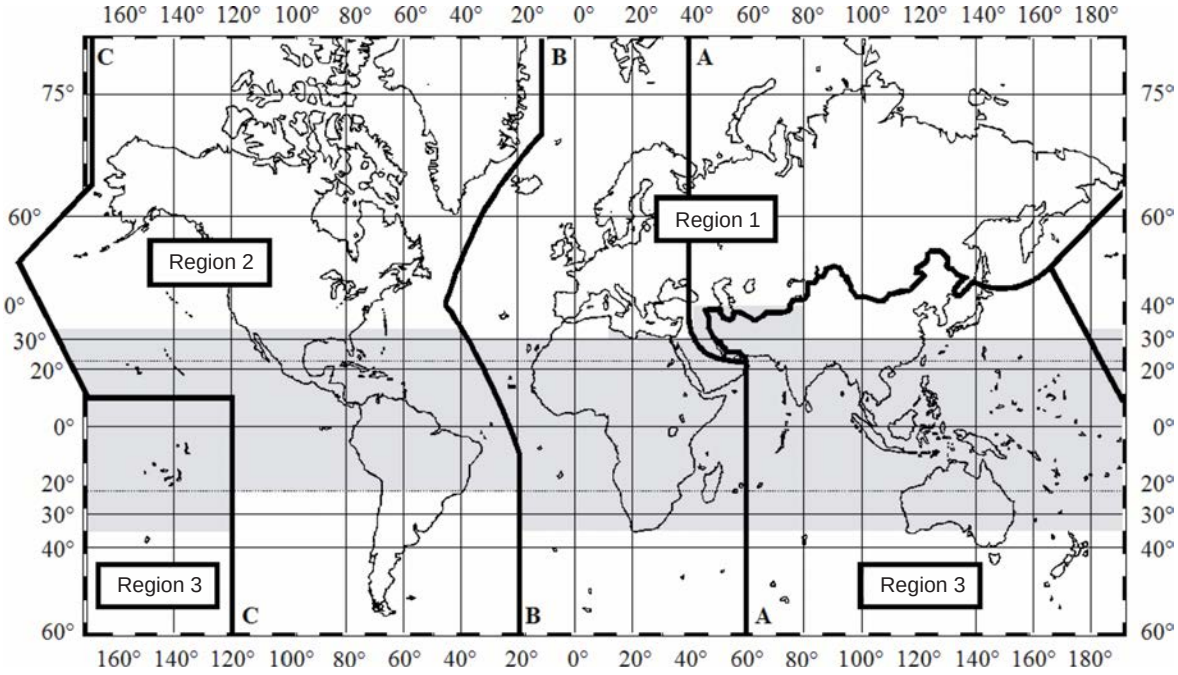


Figure 6.3.8: *The ITU regions (Source: Dodel).*

The use of decibel values thus is less likely to lead to arithmetic mistakes.

The antenna used for electromagnetic radiation or reception acts like an optical lens. The larger the antenna, the sharper the beam and the larger the gain factor of the antenna. The size of the antenna is the ratio of the antenna dimension (the diameter d if it is circular) to the wavelength of the signal it sends or receives. The **gain factor** G is

$$G = 10 \log(\mu \cdot 4 \lambda \pi^2 \cdot r^2 / \lambda^2) \text{ in dBi} \qquad (6.3.3)$$

with μ the efficiency, or approximately

$$G = 20 \cdot \log\left(2 \pi \, r / \lambda\right) \text{ in dBi} \qquad (6.3.4)$$

The **unit dBi** refers to dB of antenna gain relative to isotropic radiation with 0 dBi of antenna gain.

6.3.4.3 The Radiated Power

With the electric power P into the antenna, two aspects are important:

- The **operating point** of the amplifier. If the amplifier is not operated at saturation, the **output backoff (BO) below saturation** has to be subtracted from the value for saturation power.
- The **feed losses** V between the amplifier and antenna input, depending on the distance between the two up to $V = 2$ dB, must be taken into account.

With that information, the EIRP can be determined:

$$EIRP = P - BO - V + G \text{ (dBW)} \qquad (6.3.5)$$

It is obvious that, to generate the $EIRP$, in addition to increasing the electric transmit power P, a suitable antenna gain G can be used (the best radiofrequency amplifier still is the antenna!). Contrary to the amplifier, the antenna does not require electrical power for its operation, and, particularly, it does not age in the same way that radiofrequency amplifiers age with operating hours.

So the determining transmit parameter is the EIRP, which can arbitrarily be composed of electrical power and antenna gain. And the larger antenna not

only concentrates more sharply on the target satellite, but causes less interference for neighboring spacecraft and, in return, is less interfered with by radiation from other satellites in orbit, which are now being positioned even more closely together.

6.3.4.4 The Figure of Merit *G–T*

The **noise temperature of the receive system** t_{rec} of a ground station or satellite terminal is the sum of the sky noise t_{sky} which the antenna receives and the noise t_{sys} of the chain of amplifiers:

$$t_{rec} = t_{sky} + t_{sys} \tag{6.3.6}$$

The sky noise t_{sky} is assumed to be 40 K; the **noise temperature of the system** t_{sys} (the chain of receive amplifiers) is derived as follows:

$$t_{sys} = t_{hl} + \frac{t_{LNA}}{g_{hl}} + \frac{t_{amp}}{g_{hl} \cdot g_{LNA}} + \frac{t_{next}}{g_{hl} \cdot g_{LNA} \cdot g_{amp}} + \ldots \tag{6.3.7}$$

where:
t_{hl} = the noise temperature of the link (waveguide) between the antenna and the first amplifier,
t_{LNA} = the noise temperature of the first amplifier, the LNA,
g_{hl} = the gain of the waveguide (i.e., a number<1),
t_{amp} = the noise temperature of the following amplifier in the chain,
g_{LNA} = the gain of the LNA (the first amplifier),
t_{next} = the noise temperature of the next amplifier in the chain,
g_{amp} = the gain of the second amplifier.

A waveguide has, for example, 0.4 dB of loss per meter. The **loss figure** l_{hl} (for 0.4 dB) is

$$l_{hl} = 10^{0.4/10} = 10^{0.04} = 1.10$$

The **gain figure** g_{hl} then is

$$g_{hl} = 1 / l_{hl} = 0.91$$

The **equivalent noise temperature** t_{hl} of this component is

$$t_{hl} = (1 - 1/l_{hl}) t_0$$

where t_0 is the ambient temperature, typically 290 K, so that t_{hl} = 26.1 K.

The remaining parameters of a medium-size ground station have the following values on average:
t_{LNA} = 40 K,
t_{amp} = 600 K,
g_{LNA} = 100 (G_{hl} = 20 dB),
t_{next} = 600 K,
g_{amp} = 1000 (30 dB).

With these values, the system noise temperature of the receive system is

$$\begin{aligned} t_{sys} &= 26.1\,\text{K} + 40\,\text{K} / 0.91 + 600\,\text{K} / \left(0.91 \cdot 100\right) \\ &\quad + 600\text{K} / \left(0.91 \cdot 100 \cdot 1000\right) + \ldots \\ &= \left(26.1 + 44.0 + 6.6 + 0.0066\right)\text{K} = 76.7\text{K} \end{aligned}$$

and of the total system including sky noise temperature

$$t_{rec} = t_{sky} + t_{sys} = 40\text{ K} + 76.7\text{ K} = 116.7\text{ K or } 20.7\text{ dBK.}$$

The **figure of merit** of a ground station having, for example, an antenna gain of 48.0 dBi is then

$$G - T = (48.0 - 20.7)\text{ dBi/K} = 27.3\text{ dBi/K.}$$

6.3.4.5 Antenna Pointing

Similar to an optical lens, an electromagnetic antenna focuses sharper depending on its size, relative to the wavelength of operation. The degree of **focusing** of radiation is given relative to an isotopic radiator and is referred to as the **gain factor** *G*. In the direction of maximum focusing, the so-called boresight, the gain factor is given as a function of the size of the antenna and the wavelength λ:

$$G = G_0 = 4\pi \cdot A / \lambda^2 = 4\pi \cdot A \cdot f^2 / c^2 \tag{6.3.8}$$

where:
A = the aperture of the antenna (the circular area in the case of a circular antenna),

f = the frequency of operation ($f = c/\lambda$).

The gain maximum in boresight is called the **main lobe**. For it, the gain factor is

$$y = y_0 - p\,(x / x_0)^2, \text{also}$$

$$G = G_0 - 3\,[\theta / (\theta_0 / 2)]^2 \text{ dB} \qquad (6.3.9)$$

Outside the main beam, the upper limit of the gain is given by the approximation

$$G = 29 - 25 \cdot \log(\theta) \quad \text{in dBi} \qquad (6.3.10)$$

For angles off boresight

$$\theta > \pm 36.3^\circ \text{ results in } G = -10 \text{ dBi}$$

Figure 6.3.9 shows a typical **radiation characteristic** of a highly directive antenna.

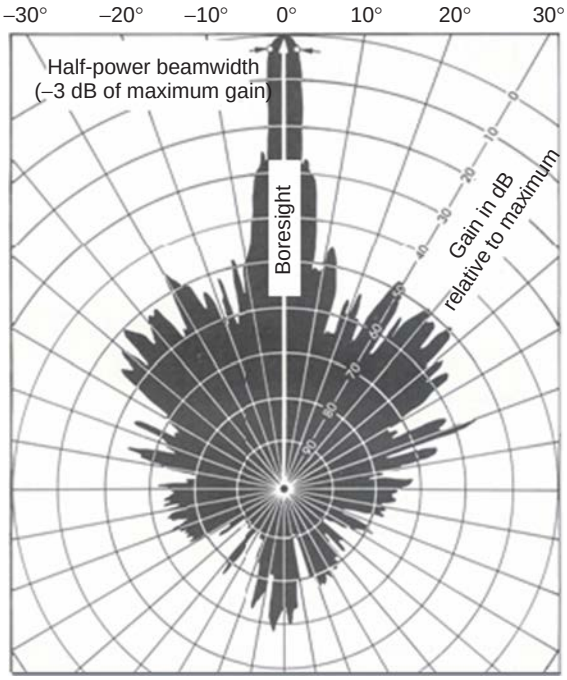


Figure 6.3.9: *Radiation pattern of a highly directive antenna (Source: Dodel).*

Antennas for small ground stations can be mounted fixed. They point at the satellite without tracking it, as long as the diurnal motion of the satellite stays within the half-power beamwidth of the antenna's main lobe. Medium-size antennas (larger than 2 m at Ku-band) are often **program tracked**. The tracker is programmed on the basis of the deterministic motion of the satellite. The alternative to program track is **step track or sequential lobing**. The power of a receive signal or a beacon is measured over a fixed period of time, say 15 seconds. Then the antenna boresight is slightly moved right or left and the measurement is repeated. If the new value is higher, the antenna boresight is moved further in that direction; if it is lower, then the pointing goes back to and beyond where it was. Following this right/left motion, the boresight is moved up and down in the same manner. In this way, it will always point right at the satellite (step track is employed, e.g., in INMARSAT ship terminals).

For the very precise **pointing** of large antennas, the monopulse system is used. The satellite signal – a pulse – is received by at least three systems, two positioned in the horizontal, two in the vertical. With the two horizontal receivers (of the three), the phase difference can be assessed, thus the misalignment of the antenna in the horizontal plane can be determined. The same is true for the two vertical receivers, allowing the misalignment in the vertical plane to be determined. Minimizing the phase difference points the antenna correctly toward the target satellite.

These antenna tracking techniques are not suited to point satellite antennas receiving signals from multiple ground stations.

6.3.4.6 Emission Dispersion and Rain Attenuation

The attenuation of the electromagnetic signals between sender and receiver is caused by radiation dispersion, attenuation in the dry atmosphere and rain attenuation.

Free Space Attenuation

The electromagnetic power radiated from a point in space is distributed uniformly in space on the surface of a sphere whose center is the source and whose radius is the distance from the source. The power is

dispersed from the point source to the surface of the sphere by the factor

$$fd = (4\,\pi \cdot d \cdot f / c)^2 \qquad (6.3.11)$$

In dB, this **free space attenuation** is

$$FD = 10 \cdot \log(4\,\pi \cdot d \cdot f / c)^2 = 20 \cdot \log(d \cdot f) + 92.44 \text{ dB}$$

with d in km, f in GHz.

The difference in the free space attenuation FD values for a geostationary satellite between the nadir value and the value in the direction toward the edge of the Earth is

$$FD_{\text{edge of earth}} = FD_{\text{nadir}} + 1.4 \text{ dB} \qquad (6.3.12)$$

Atmospheric Attenuation

The attenuation of the dry atmosphere AD adds to the free space attenuation. Typical values for frequencies from 1 to 30 GHz are summarized in Table 6.3.3.

Rain Attenuation

Rain attenuation adds to the attenuation in the atmosphere and the signal dispersion in space, see Figure 6.3.10.

Figure 6.3.10 also depicts the **ionospheric effect** on propagation, which increases for smaller frequencies while the rain attenuation increases for increasing frequencies; the minimum of ionospheric and rain attenuation occurs at 5 GHz. Local maxima of the rain attenuation lie at 24 GHz, the molecular attenuation of hydrogen, and at 60 GHz, the molecular attenuation of ozone. Interestingly, at around 100 GHz rain attenuation no longer increases but stagnates.

Path Loss

The path loss PL is the sum of free space attenuation, atmospheric attenuation and rain attenuation:

$$PL = FD + AD + RD + LM \qquad (6.3.13)$$

where LM is a sufficient link margin (at least 1 dB).

G–T Degradation Due to Rain

Since the temperature t_{sky} goes into the figure of merit G–T of the ground station and increases during rain, the G-T degrades by the value $\Delta g/t$:

$$\Delta g / t = [t_{sys} + t_{sky}\ (1 - 10^{-AA/10})] / t_{sys} \qquad (6.3.14)$$

where:

AA = Additional Attenuation: rain attenuation (in dB),

t_{sys} = system noise temperature (in K),

t_{sky} = 40 K.

Since the transmission must function during rain, the figure of merit of the (dry) ground station must be corrected by this factor $\Delta g/t$.

6.3.4.7 Interference

The **electromagnetic compatibility of radio systems**, particularly of satellite systems, is regulated through the ITU Radio Regulations (RR). These regulations

Table 6.3.3: *Atmospheric attenuation for higher frequencies.*

Elevation	1 GHz (dB)	4 GHz (dB)	6 GHz (dB)	11 GHz (dB)	14 GHz (dB)	20 GHz (dB)	30 GHz (dB)	45 GHz (dB)
0°	0.61	1.54	1.90	3.02	4.67	14.74		
1°	0.60	1.53	1.87	2.79	4.01	9.01	13.25	
5°	0.22	0.55	0.67	0.99	1.42	3.25	4.63	7.01
10°	0.11	0.29	0.35	0.52	0.74	1.70	2.42	2.91
15°	0.08	0.19	0.24	0.35	0.50	1.15	1.63	1.73
25°	0.05	0.12	0.15	0.21	0.31	0.71	1.01	1.43
45°	0.03	0.07	0.09	0.13	0.18	0.41	0.59	0.73
60°	0.02	0.06	0.07	0.11	0.15	0.42	0.49	0.64
90°	0.02	0.05	0.06	0.09	0.13	0.30	0.42	0.59

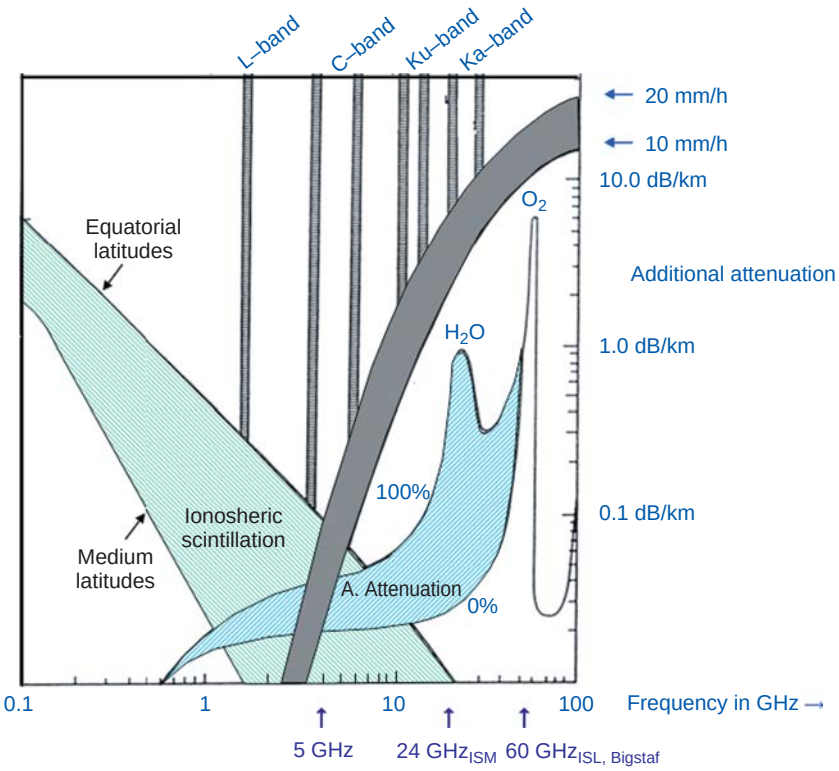


Figure 6.3.10: *Rain attenuation versus frequency (Source: Messerschmid).*

define the relevant satellite and ground station parameters and limit the radio interference. The different cases of radio interference are portrayed in Figure 6.3.11.

The own signal power C and the interference power I, both in dBW, lead to the **carrier-to-interference ratio**

$$\begin{aligned} C-I &= EIRP_{\text{interfering sat}} - EIRP_{\text{own sat}} \\ &\quad -\Delta G_{\text{sat}}(\varphi) - \Delta G_{\text{ground station}}(\phi) \\ &= EIRP_{\text{interfering ground station}} - EIRP_{\text{own ground station}} \\ &\quad -\Delta G_{\text{ground station}}(\varphi) - \Delta G_{\text{sat}}(\phi) \end{aligned} \qquad (6.3.15)$$

and C–$I > 35$ dB for one case of interference, or C–$I > 30$ dB for all interferers in the system. New systems are allowed only if they prove not to cause interference. The computation of the level of interference I is defined in the ITU Radio Regulations in RR Appendix 29. It can be referred to either the own carrier power C so that, for example, C–$I > 35$ dB, or to the own

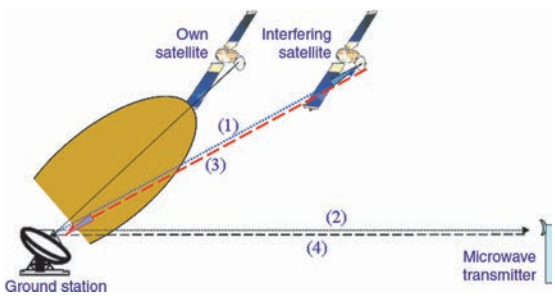


Figure 6.3.11: *The interference between ground station, satellite and microwave links: (1) the ground station interferes with a neighboring satellite; (2) the ground station interferes with a terrestrial microwave link; (3) the ground station is interfered with by a satellite; (4) the ground station is interfered with by a terrestrial microwave link (Source: Dodel).*

noise power N. The first concept is the C–I method, the second the $\Delta t/t$ method.

The $\Delta t/t < 6\%$ Method

The grave disadvantage of the C–I method is that it does not consider how good or bad the interfered C–N

is. If it is high (>10 dB), the victim suffers greatly and it is reasonable to ask for $C–I > 30$ dB. If it is relatively low ($C–N < 10$ dB), the harm to the carrier is less. So it is not reasonable to ask globally for $C–I > 30$ dB. The ratio $C–I$ is therefore referred to the $C–N$ of the interfered carrier:

$$(C-I)-(C-N)>12.2 \text{ dB}$$

The $C–I$ must be 12.2 dB better than the $C–N$; the interferer degrades the carrier-to-noise ratio by < 0.25 dB. Cancelling the carrier power C leads to $N–I > 12.2$ dB (with N the thermal noise, I the interference). Cancelling the bandwidths B and the factor K (remember that $n = t \cdot b \cdot k$) in $N–I$, we get $T - T_i$ or for $T_i \equiv \Delta T$ the interference, or $T - \Delta T > 12.2$ dB, or equation.

$$t/\Delta t > 16.7 \quad \text{or} \quad \Delta t/t < 6\,\%$$

The interference is not referenced to the carrier power but to the noise power. The value $\Delta t/t < 6\%$ corresponds to a $C–I = 30$ dB for the carrier-to-noise ratio $C–N = 17.8$ dB, a relatively high value. Typical $C–N$ values in digital transmission including error correction are in the range of 5 to 10 dB.

Relating the interference to the thermal noise of the victim carrier makes the interference independent of the error correction employed on the victim carrier. With an **error correction** which reduces the $C–N$ of the transmission by, say, 5 dB, the error correction does not change the thermal noise N (or T), but only reduces the required carrier power C.

The $C–I > 35$ dB method, in this case, would allow the interference 5 dB more power, which would not be in the interest of whoever invests in coding. The $\Delta t/t < 6\%$ method, in contrast, can be applied without regard to the possible coding that may be used, since it is related to the noise temperature T.

Thus, $\Delta t/t < 0.06$ or $< 6\%$ or < 0.25 dB; however, the interference amounting to 6% applies per case of interference! If there are 10 cases acting on the victim carrier, then the total interference is 60% so that noise plus interference becomes 160% (or $10 \cdot \log(1.60) = 2.2$ dB) of the noise.

Uplink and Downlink

For serial links, such as the satellite uplink and downlink, the **individual values** of $\Delta t/t$ are added up similar to carrier-to-noise ratios:

$$\sum \Delta t/t = \Delta t_{up}/t_{up} + \Delta t_{down}/t_{down} \qquad (6.3.16)$$

The interference-to-noise ratio of the uplink is added to the interference-to-noise ratio of the downlink.

The Interference Power

To determine the interference power, both the **carrier power** C (in dBW) as well as the **spectral power density** C_0 (in dBW/Hz) are used. The value of C_0 in the center of the spectrum is used, which is higher than the average over the bandwidth b by a factor F:

$$C_0 = C - 10 \cdot \log(b) - F \qquad (6.3.17)$$

where F in the case of FM, for instance, would be 8.5 dB. For PSK, F is smaller since the power of the carrier modulated digitally is spread much flatter over the bandwidth B.

For the carrier-to-noise $C–N < 17.8$ dB the requirement $\Sigma\Delta t/t < 6\%$ allows a higher interference power (an extended coexistence under conditions of interference) than does the $C–I > 30$ dB method, without undue impact on the transmission.

The requirement $C–N > 17.8$ dB would lead to lower interference powers for $\Sigma\Delta t/t < 6\%$ than with $C–I > 30$ dB. Such $C–N$ values, however, are not common.

The $\Delta t/t$ method adapts better to the $C–N$ of the interfered-with transmission than does the requirement $C–I > 30$ dB, which dates back to analog transmissions.

The Calculation

The carrier-to-interference ratio $C–I$ (in dB) is determined by

$$C - I_d = P_{si} + G_{sti} + G_{esi} - L_{di} - \left[P_{sj} + G_{stj}(\theta_{ji}) + G_{eri}(\theta_{ji}) - L_{dj} \right] \quad \text{in dB} \qquad (6.3.18)$$

where:
P = transmit power in dBW,
G = antenna gain = $29 - 25 \cdot \log(\theta)$ in dBi and θ in °,
L = free space attenuation in dB,
e = Earth station,
s = satellite,

u = uplink,
d = downlink,
r = receive,
i = interfered-with system,
t = transmit,
j = interferer.
The **sum carrier-to-interference ratio** C–I is

$$\frac{c}{i_{\text{total}}} = \frac{1}{\frac{i_{\text{up}}}{c} + \frac{i_{\text{down}}}{c}} \qquad (6.3.19)$$

$$\text{and } C - I = 10 \cdot \log\left(c\right) \text{in dB}.$$

For similar links with $L_{\text{ui}} = L_{\text{uj}}$ and similar transmit powers $P_{\text{ei}} = P_{\text{ej}}$, with an angular separation of, for instance, 3°, the result is $G = 29 - 25 \log(\theta) = 20.1$ dB. For a nondiscriminating satellite antenna (receiving the interferer with the same gain G as the victim) the result is

$$\begin{aligned} C - I_{\text{u}} &= G_{\text{eti}} + G_{\text{rsi}} - G_{\text{etj}}\left(\theta_{\text{ju}}\right) - G_{\text{sri}}\left(\theta_{\text{ji}}\right) \\ &= 40.0 + 33.5 - 20.1 - 33.5 = 19.9 \text{ dB}. \end{aligned}$$

For similar uplinks and downlinks, C–I_{d} = 19.9 dB and C–I_{total} = 16.9 dB. With one interfering satellite to the right and another one to the left, each with similar power, the sum is C–I_{total} = 13.9 dB.

It could be worse: with an interference power of 6% of the thermal noise, the C–I = 12.2 dB per link and interferer (i.e., 9.2 dB uplink and downlink together) and $10 \cdot \log(n)$ dB less when there are n interferers.

Satellite systems today are beginning to be **interference limited**; the disturbance by interference reaches or exceeds the disturbance by thermal noise plus intermodulation.

The link budget with thermal noise and intermodulation is described in the following section.

6.3.4.8 The Link Budget

The **carrier levels** of the transmission must be dimensioned to yield the desired **service quality**. Neither the noise power n nor the carrier power c are sufficient for the purpose, but rather the **ratio** c/n or, in dB, C–N:

$$\begin{aligned} \text{where } C &= 10 \cdot \log(c) \quad \text{in dBW} \\ &\left(\text{or. } c = 10^{0.1 \cdot C} \text{ in W}\right) \\ \text{and } N &= 10 \cdot \log(n) \quad \text{in dBW} \\ &\left(\text{or. } n = 10^{0.1 \cdot N} \text{ in W}\right) \\ &= T + K + B \quad \text{in dBW, noise power according to Johnson} \\ &= N_0 + B \end{aligned}$$

where:
T = noise temperature in dBK,
K = Boltzmann's constant = −228.6 dBJ/K,
N_0 = noise power density in dBW/Hz,
B = bandwidth in dBHz.

The following terms have proven useful:

- C–$N = C - N$ in dB, or $c/n = 10^{0.1 \cdot C/N}$, the carrier-to-noise ratio at RF
- $EIRP = P + G - V$ in dBW, the **e**quivalent to isotropic radiated power
- $PL = FD + AD + RD$ in dB, the conglomerate path loss
- G–$T = G - T$ in dBi/K, the figure of merit of the receive system.

With these,

$$\underset{[\text{dB}]}{C - N} = \underset{[\text{dBW}]}{EIRP} - \underset{[\text{dB}]}{PL} + \underset{[\text{dBi/K}]}{G - T} - \underset{[\text{dBWs/K}]}{K} - \underset{[\text{dBHz}]}{B} \qquad (6.3.20)$$

or

$$\underset{[\text{dB}]}{C - N_0} = \underset{[\text{dBW}]}{EIRP} - \underset{[\text{dB}]}{PL} + \underset{[\text{dBi/K}]}{G - T} - \underset{[\text{dBWs/K}]}{K} \qquad (6.3.21)$$

This is the credit side of the link budget. The debit side is given by the energy per bit E_{b} relative to the noise power density N_0 required for a given **quality of transmission**, expressed in the BER of, for example, 10^{-5}. With these,

$$\begin{aligned} C - N_0 &= E_{\text{b}} + \text{bit rate} - N_0 - CG + IM \\ &= E_b - N_0 + \text{bit rate} - CG + IM \\ &= \underset{[\text{dB}]}{E_b - N_0} + \underset{[\text{dBHz}]}{10 \cdot \log(\text{bit rate})} - \underset{[\text{dB}]}{CG} + \underset{[\text{dB}]}{IM} \end{aligned} \qquad (6.3.22)$$

where CG is the coding gain that can be subtracted from the C–N, and IM is the implementation margin, which must be added to it. For the balance,

$$C - N_{0.\,\text{credit}} \geq C - N_{0.\,\text{deficit}}$$

must be given.

The Total Link via Satellite

The **noise contribution of the total link** comprises the thermal noise n_{uplink} of the uplink, the radio interference i_{uplink} of the uplink, the intermodulation noise n_{intermd} in the satellite transponder, the thermal noise n_{downlink} of the downlink and the radio interference i_{downlink} of the downlink caused by satellites and terrestrial sources.

To determine the sum of these contributions, the noise powers related to the carrier power must be added:

$$n_{\text{credit}}/c = n_{\text{uplink}}/c + i_{\text{uplink}}/c + n_{\text{intermd}}/c + n_{\text{downlink}}/c + i_{\text{downlink}}/c \quad (6.3.23)$$

With n_{credit}/c then $c_{\text{credit}}/n = 1/(n_{\text{credit}}/c)$ can be obtained by inversion, and

$$C - N_{\text{credit}} = 10 \cdot \log\left(c_{\text{credit}}/n\right) \quad \text{in dB} \quad (6.3.24)$$

The Link Margin

Finally,

$$C - N_{0.\text{credit}} = C - N_{0.\text{deficit}} + R \quad (6.3.25)$$

where R is the link margin (it should be at least 1 dB).

6.3.5 Ground Station Operation

The ground station, as the link between the control center and the space vehicle, must be controlled and monitored. This is carried out by the **monitoring and control system**, operated by personnel who configure the ground station and establish and maintain the link to the space vehicle, according to well-defined procedures. The operation of this link depends on the performance of the ground station, the space vehicle and its orbit (LEO, GEO, deep space, LEOP).

6.3.5.1 Command and Control System

The systems of the ground stations employed for the operation are configured and continuously observed using the monitoring and control system (M&C system). To guarantee secure operation and fast reaction times, the following features of the M&C system (Figure 6.3.12) are emphasized:

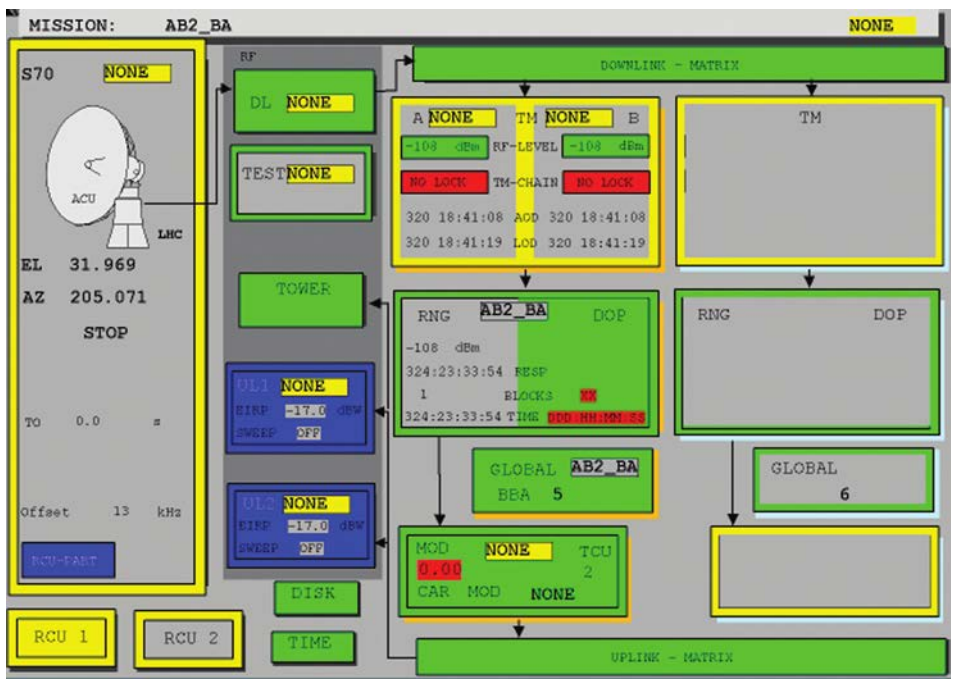


Figure 6.3.12: *M & C system at the ground station (Source: DLR).*

- Clear structure and logical levels of detail (see the example below).
- Information through color coding and acoustic signals (especially in the case of a fault!).
- Display of important parameters indicating the quality of the link, such as antenna AGC and azimuth and elevation angles.
- High update rates for M&C system parameter visualization.
- Logging of the control inputs and equipment responses.
- A stable computer system.

6.3.5.2 The Operation of a Pass

The operation of a pass (i.e., a flyover of the satellite above the ground station) starts with an entry in the **schedule**, the order book of the ground station, listing which mission is to be supported at what time, using which antenna. Each operation during a pass is executed according to the procedure, and each action and event is logged. The operators are in contact with the control center during the entire pass.

The operation of the pass is divided into three parts, as follows.

Pre-Pass (Preparation of the Pass)
The ground station is being prepared for the pass. All systems are configured and loaded with the current orbit data. Using the **data flow test** (DFT), the receive function of the station is tested with simulated telemetry data. In a **command test** the correct configuration of the uplink is verified. Then the antenna is pointed in the proper direction to start communication with the satellite as soon as it has reached a sufficient elevation.

The Pass (Execution of the Pass)
The antenna can be **tracked** differently during the pass, in autotrack, predict track or step track mode. The link with the satellite can be established after **acquisition of the satellite signal**, possibly after a sweep. Commands can be sent to the space vehicle and telemetry received. High-rate data can be stored in the ground station(s) for distribution to the control center and to users. Navigation data (angle values, Doppler data, ranging data) is stored according to need. The planned end of the pass is at the predicted end of the contact with the spacecraft or, sometimes, the pass is terminated with the premature loss of the signal (loss of lock) by the ground station receiver.

Post-Pass Processing
After the pass the antenna is configured to a defined idle position. The payload and navigation data is distributed, and a **post-pass report** containing all relevant data concerning the pass, as well as a failure report, if necessary, recording and explaining experienced anomalies, is prepared.

6.4 Operations for Human Space Flight

Dieter Sabath and Thomas Kuch

In comparison to unmanned missions, human space flight missions offer great advantages. The crew onboard is able to work and solve **scientific problems** directly: that is, they can observe in situ; intervene directly if the need arises and pursue defined goals; perform experiments; and solve problems in case of anomalies. Nevertheless, a human presence on-board requires considerably improved **safety precautions** and therefore such missions need longer preparation times, more redundancies, more thorough reviews for documents and tools, and thus offer reduced flexibility for short-term changes.

Since 2000 Europe has used the **International Space Station** (ISS) for human space flight activities. For the transportation of astronauts the US Space Shuttle or the Russian Soyuz capsules are used under the responsibility of the respective space agencies, NASA and Roscosmos. With the berthing of the European **Columbus module** in February 2008 (Figure 6.4.1), ESA assumed responsibility for operating the module at the ISS. The **Columbus Control Center** (Col-CC), operated by the German Space Operations Center (GSOC) at DLR in Oberpfaffenhofen, close to Munich, is monitoring and commanding the module under ESA responsibility.



Figure 6.4.1: *Assembly status of the ISS in spring of 2008 with the Columbus module docked to Node 2* Harmony *(Source: NASA).*

The Col-CC closely cooperates with the User Support and Operations Centers (USOCs) dedicated to payload operations in nine European countries, the Engineering Support Centers (ESCs) in Bremen and Turin, as well as the control centers of the international partners at the Johnson Space Center (JSC) in Houston, Texas, at the Marshall Space Flight Center (MSFC) in Huntsville, Alabama, at the TsUP (Tsentr Upravleniya Polyotom, Flight Control Center) in Moscow, and at the Space Station Integration and Promotion Center (SSIPC) in Tsukuba, Japan (see Figure 6.4.2). Communication with the Japanese control center is via NASA communication lines.

The European user centers (see Figure 6.4.2 and Table 6.4.2) are responsible for executing the experiments in the Columbus module or in other ISS modules, coordinated by the Col-CC. NASA JSC is responsible for the US segment of the ISS and has overall responsibility for ISS operations. The second NASA control center in Huntsville coordinates the operations of the US payload racks and experiments. The ZUP controls the operations of the Russian segment whereas the SSIPC is responsible for the Japanese module.

6.4.1 Preparation

Preparing a human space mission is a demanding and complex job because of the high responsibility for the crew on-board. Due to their high costs and complexity, human missions are mostly planned and conducted by **international cooperation**. This leads, in contrast to most satellite missions, to a variety of **interfaces** between the partners which have to be respected in the design. The US **Constellation Program**, currently under development, is an exception. This program is to lead to a return to the Moon by the USA using the Ares-1 and Ares-5 launchers as well as the Orion capsule as national assets.

During the preparation phase of a mission the **operations concept**, the **procedures** and the **tools** have to be determined, designed, developed, tested and released. This is done partly in parallel with the development of the flight hardware and demands close coordination between the development and operations teams. One approach is to integrate operations team members in the development team and vice versa to ensure an exchange of experience and data. This allows operational aspects to be taken into account during development and to provide the expertise of the development team to the operations team.

This approach was partly introduced during the Columbus operations preparation phase. Additionally, the operations team benefits from the experience of the American and Russian colleagues, both with long-standing experience in human space operations. The Columbus module is operated by the Col-CC, which has three major tasks: first, to monitor the **orbital system** (the Columbus module), second to coordinate the European **experiments and payloads** on the ISS, and third to operate the **ground communication infrastructure** with the communication connections (lines) to all partners. For the first two tasks the Columbus flight control team (FCT) has two mission control rooms, a planning room and several user rooms, as well as a backup mission control room in a physically separate building which can be used in case of catastrophic events (fire, aircraft impact). One control room (see Figure 6.2.3) is used for preparing a mission increment while the other mission control room accommodates the team for the real time operations execution. The control rooms are equipped with up to 14 consoles displaying the telemetry data of the Columbus module in various formats to the FCT and allow access to the data and video channels of the ISS.

The FCT consists of flight controllers with dedicated task assignments. During the mission the so-called console positions are coordinated by a flight director, who is responsible for the overall control of

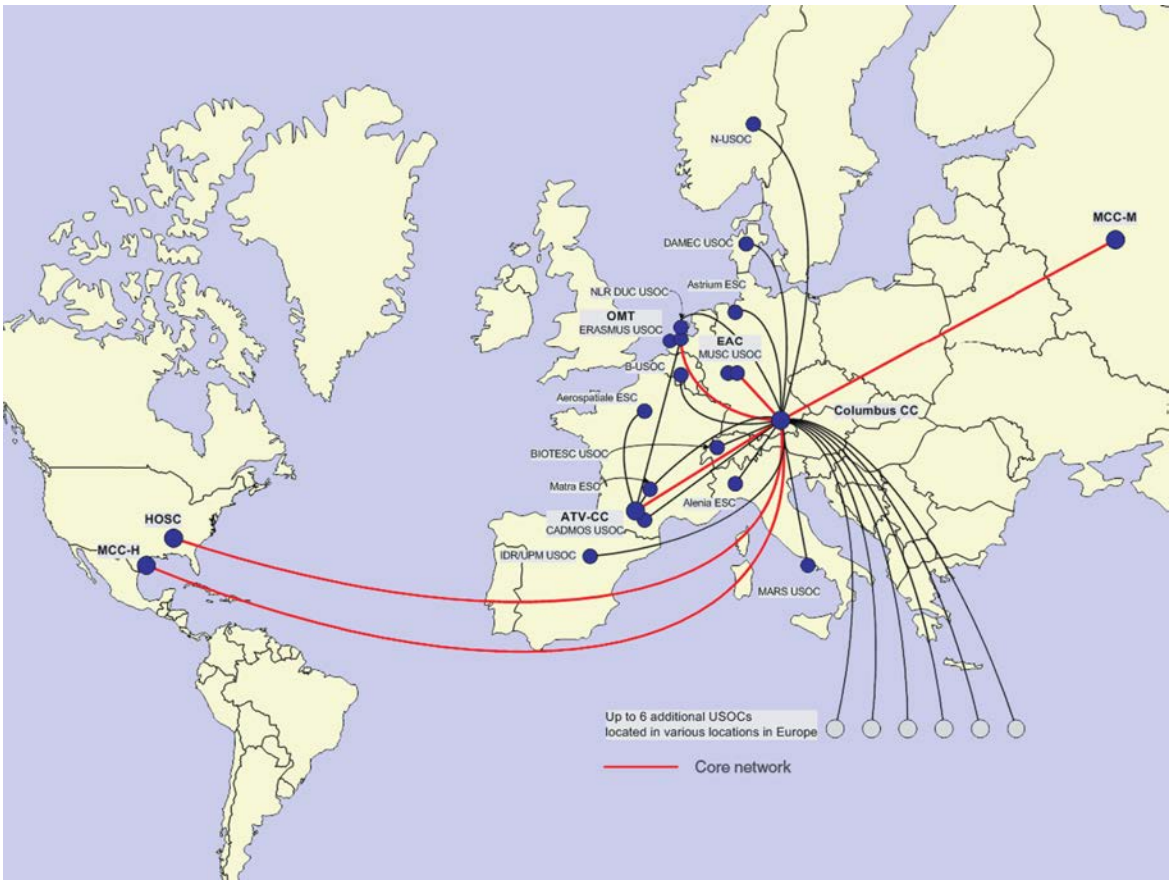


Figure 6.4.2: *The European ground communication network (Source: DLR).*

operations. The flight director collects the relevant information from all team members and decides on the further course of action. The flight director is in direct contact with the flight directors in Houston and Moscow and with the ground control team (GCT) at Col-CC, and informs the ESA managers about all events in the Columbus module. Additionally, the flight director coordinates decisions about crew activities together with the crew surgeons and biomedical engineers (see Section 6.4.4.2).

The most important control room **positions in the control room** of the Col-CC are described in Table 6.4.1. The consoles are also available for the main task of the FCT commanding the Columbus module. The monitoring and commanding of the module relieve the on-board crew from routine tasks, enabling the crew to concentrate on the execution of experiments and maintenance tasks on the ISS which cannot be performed from the ground.

The FCTs in the control centers rely on the support of the ground control teams who are responsible for the communications lines and the configuration of the control center subsystems. The ground control team of the Col-CC supports all human missions from a dedicated control room (see Figure 6.2.4) and is in close contact with the ground control teams at the European user centers as well as at the control centers of the international partners.

6.4.1.1 Tool Development

The **tools** of the FCT consist of on-board procedures, interface procedures, documents, telemetry displays

Table 6.4.1: *Positions of the flight control team.*

Position	Task
Columbus Flight Director (FD)	Head of the FCT
Operations Coordinator (OC)	Coordination of the USOCs and their experiments. Deputy to the FD
Col-System Engineer	Monitoring and commanding of the ECLSS, the power subsystems and the thermal subsystem
Data Management System (DMS) Engineer	Responsible for the data management system collecting data from subsystems and experiments
Communications Officer ("Comms")	Monitoring and commanding data transfers from and to the ISS
Columbus Operations Planner (COP)	Monitoring and replanning all European activities on the ISS
Crew Interface Coordinator (Eurocom)	Communication with the ISS crew for European subsystems and experiments

and support software. These tools are developed either by the FCT or in close cooperation with the flight controllers.

For monitoring and commanding the Columbus module, a **monitoring and control system** (MCS) is used. The MCS processes the data of the Columbus module and displays the results to the flight controller using graphical and alphanumeric displays. The displays are designed by the FCT and are upgraded according to experience gained during simulations and operations. Figure 6.4.3 shows an example of an alphanumeric parameter display with simulated data for the Columbus module.

Among others, these parameters consist of temperature, pressure, humidity and atmospheric composition measurements in the Columbus module as well as the measurement of the power consumption, the thermal parameters and the status of the communications lines. These parameters are mainly used for monitoring the Columbus module and are permanently observed by the FCT. To ease the control tasks of the FCT and to ensure the highest possible failure detection rate, the parameters are checked against dedicated "soft" and "hard" boundaries (limits). The soft boundary is linked to a "caution" while the hard boundary is associated with a "warning." This is indicated to the flight controller by a color change in the display of the parameter, for example yellow for "caution" and red for "warning." In case of serious failures a station-wide **caution and warning system** (CWS) is activated to inform the astronauts of such anomalies or failures.

Besides the parameter displays, the flight controllers can also use graphical displays, for example to show the current position of the ISS in its orbit around the Earth or the current attitude of the ISS with respect to its orbit (Figure 6.4.4).

In addition, the beginning and end of upcoming **communication slots** with the ISS, the on-board short-term plan (OSTPV) of the astronauts and cosmonauts, as well as a list of the current cautions and warnings, are displayed.

The interface procedures for the operations of the control center, the communication between the flight controllers, as well as those between the different control centers, have been specified in several documents, agreed on by the flight and ground control teams together with European and international partners, and tested in simulations. To ensure the safety and correctness of the products and documents they are validated in a multistep approach. The validation of the on-board procedures, tools and interface procedures starts with stand-alone testing, followed by simulations which become evermore complex and finally include anomalies or failures. After successful validation the products and tools can be used in operations.

6.4.1.2 Training, Simulations and Certification

Besides the tools and documents which have to be validated for operations, the personnel also have to be trained and certified to work on consoles. The **training and certification plan** is adapted to the various needs of the dedicated control center; the approach at the Col-CC is described as follows.

During **baseline training** the use of the tools needed when working on consoles has to be learned, for example the use of the voice communication system and the monitoring and control system. Additionally, new team members must attend courses to learn the internal procedures of the control center, the



Figure 6.4.3: *Alphanumeric display of selected parameters of the Columbus module (Source: DLR).*

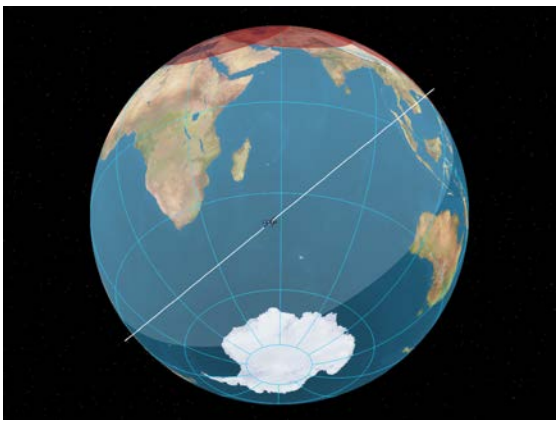


Figure 6.4.4: *ISS orbit display (Source: DLR).*

functioning of the Columbus module and its payloads, as well as of the ISS in general.

After this first step the candidate flight controller takes part in first **simulations** which are set up at the control center. During these simulations the behavior of the Columbus module is modeled by a computer and the reaction of the subsystems to the commands from the FCT is presented on the screen. The goal of the simulations is to improve the teamwork of the flight controllers under realistic conditions and to present failure cases of the Columbus module to the team. The reaction and the problem-solving approach of the team and of each member are monitored and assessed.

After participation in internal simulations the new flight controller can take part in simulations with other European centers. During these simulations more complex procedures with several partner control centers are executed. The last step before final certification for a mission increment is successful participation in simulations with NASA control centers, because Col-CC has the closest relationship to this international partner. For payload operations in the Russian segment, there are also simulations with the Russian control center.

6.4.1.3 Interface Coordination with the International Partners

One of the most important tasks during mission preparation is the coordination between the international

partners. The Columbus module is one of the habitable modules of the ISS. Columbus is directly attached to the US segment and receives its resources (air, power, thermal and communication services) from this segment. Hence, the **interfaces** with the necessary subsystems of the ISS and with the adjacent modules have to be described in detail. On the one hand this has already been accomplished during the development phase of the modules when the mechanical, electrical, hydraulic and communication interfaces between the modules were defined; on the other hand, there are operations interfaces which have to be agreed by the FCTs in Houston, Moscow, Tsukuba and Oberpfaffenhofen.

To achieve these agreements in time, **common panels** among the FCTs are established, to coordinate a joint approach during docking, activation and commissioning as well as during steady-state operations. In the panels interface procedures between the control centers are determined and access to the necessary tools (Web pages, documents and software tools) of the partners is agreed. Additionally, cooperation is trained in common simulations (see Section 6.4.1.2). Starting in 2001 the Columbus launch and activation was prepared in the **European Joint Operations Panel** (EJOP), which consisted of teams from Houston and Munich. In the year before the launch the teams were increased by the Space Shuttle FCT, representatives of the shuttle crew and, as necessary, by teams from the other international partners (e.g., a Japanese team for interfaces with the Kibo module).

A similar approach was chosen for the preparation of joint missions during the assembly phase of the ISS, for instance during the ESA Astrolab mission in 2006. In this case representatives of the FCTs from Moscow, Houston, Huntsville and Munich met several times to finalize the interface procedures needed for the mission and agree on the use of resources on the ground and on the ISS.

In all cases the preparation team is almost identical to the operations teams working on consoles during the mission. To certify them for the mission, the teams were trained in a staggered approach using operations simulations which increased in complexity.

6.4.1.4 Planning

The time available to the astronauts to perform all the necessary tasks on-board the ISS and therefore to execute experiments is restricted and very expensive. To ensure that it is used to the maximum extent possible, all activities on-board the ISS as well as the related ground activities are preplanned and coordinated. It is the task of the planning teams in Europe and at the partner control centers to establish these plans.

The planning of the activities to be performed on-board is done in half-year slices called **increments**. An increment starts with the launch of a new core team to the ISS and ends with the landing of the same team about six months later.

In the first phase, Phase A (see Figure 6.4.5), an **on-orbit summary** (OOS) is established, describing the task for the entire six-month increment period in an executable sequence of events. This preplanning ensures that the time available for each activity of the astronauts and cosmonauts is sufficient to perform all tasks and that all necessary resources are available in time. Additionally, the preplanning determines which tools, devices, experiments and disposable matter have to be provided on-board and ensures that the crew on-board and on the ground are trained for their tasks.

The next step, Phase B, is the **weekly look-ahead plan** (WLP), which is always established 10 days in advance. The WLP details the activities for each day of the week, showing all required tools and procedures. Based on the WLP the operations teams and the crew can coordinate their preparatory activities.

The last step in the planning approach, Phase C, is the generation of the **on-orbit short-term plan** (OSTP), released one week in advance. This plan describes the sequence of activities of the crew on-board the ISS and on the ground in the greatest detail, including all procedures and tools and all changes compared to the WLP. The OSTP is the basis for the daily planning conference the evening before, when the tasks for the following day are coordinated with the astronauts. The scheduled tasks are performed following the plan as closely as possible (Figure 6.4.5). The plan can be modified and adapted for the required changes caused by anomalies that may occur when an experiment is executed, or by other unforeseen events. The planning team tries to shift delayed activities to one of the next free slots in the schedule.

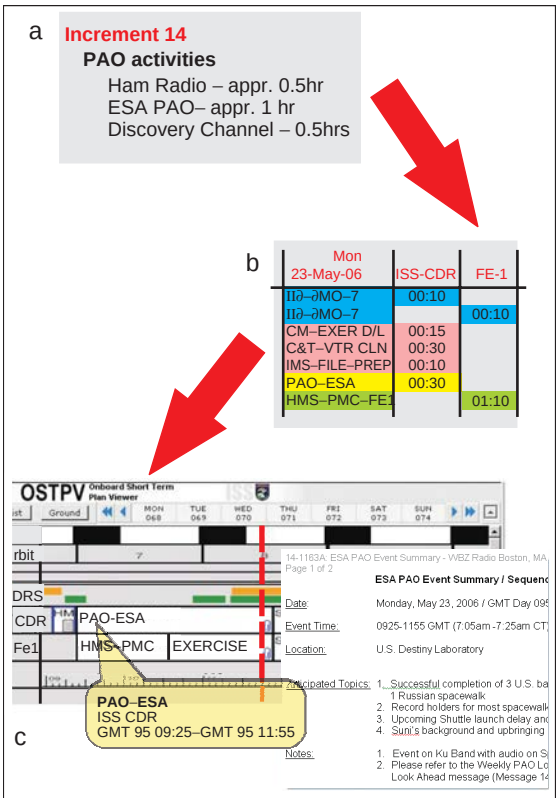


Figure 6.4.5: *Planning steps (Source: DLR).*

6.4.2 System Operations of an ISS Module

The ISS is dedicated to performing research under microgravity conditions. Nevertheless, a substantial part of the available time has to be invested in securing and maintaining the necessary living and working conditions for the astronauts. Two of the three astronauts on-board the ISS in 2007 had the sole task of **performing maintenance** on the ISS modules and **repairing faulty parts**. Only half of the working time of the third crew member was available for conducting experiments and for payload operations.

In addition to the work of the ISS crew, the FCTs in the control centers provide an enormous amount of support to keep the ISS operational. In the control centers in Houston and Moscow, and since 2008 also in Munich and Tokyo, the FCTs work around the clock to operate the module assigned to their responsibility. The teams ensure that the module is working correctly, that the necessary maintenance is performed in time, that the environmental conditions for the astronauts are met, and that the payload operations can be performed.

6.4.2.1 System Operations of the Columbus Module

The main task of the FCT at the Col-CC is to activate, command and monitor the Columbus module after its berthing with the space station (see Figure 6.4.6).

The FCT, which is available 24 hours a day in three shifts during steady-state operations, has to ensure that the **environmental control and life support system** (ECLSS) is working correctly, and that the **power supply** and distribution are working properly and providing sufficient power to the subsystems and experiments. Additionally, the thermal subsystem for **cooling** the subsystems and experiments has to be monitored, and it must be assured that all **data, video and audio connections** are properly routed. In case of anomalies or faulty equipment the team has to perform countermeasures and provide a description of the failure to the **engineering support team** for analysis.

The engineering support team consists of experts from the manufacturers of the various subsystems of



Figure 6.4.6: *Columbus module at ISS Node 2 (*Harmony *module) (Source: ESA).*

the Columbus module. Most of the team members reside at the ESCs in Bremen and Turin, while a small team supports the FCT at Col-CC directly. They are informed by the FCT about failures in the module and provided with the necessary data. The ESC analyzes the anomaly, tries to reproduce the failure in simulators on the ground, and issues a recommendation on how to solve the problem. It is the task of the FCT to execute the recommended procedures in order to regain the full functionality of the module.

6.4.2.2 Environmental Control and Life Support System

The Columbus module is one of the pressurized elements of the ISS and has to provide a **shirtsleeve environment** for the astronauts. Despite the lack of its own resources for power, air and cooling (these are provided via the US elements of the ISS), the FCT is responsible for the correct environmental conditions inside the module. The operator on the systems position (see Table 6.4.1) monitors the ECLSS and the parameters for the environmental conditions in the module. It has to be ensured that the air inlet and outlet flows, the air circulation within the module, the oxygen and the CO_2 concentrations, as well as the humidity, are in the correct range.

It is also the task of this operator to perform preventive maintenance of the ECLSS together with the ISS crew. In the case of anomalies the operator will initiate the commands to switch over to the backup systems. The monitoring and the maintenance of the smoke detectors are also under the responsibility of the operator on the systems position.

6.4.2.3 Power Supply and Thermal Control

There are two **independent power connections** providing power from the solar generators, or in shadow phases from the batteries, to the ISS modules. Hence, close cooperation between the responsible positions in the various control centers is necessary, because increased power consumption in one module or the partial failure of a power subsystem directly influences all modules.

Monitoring the power subsystem within the Columbus module and coordination with the ISS partners are also the responsibility of the operator on the systems position. The flight controller has to ensure that the subsystems and experiments receive the power needed without exceeding the maximum allowed power consumption. Since a large part of the energy is transformed into heat, the experiments and subsystems have to be **actively cooled**. The excess heat is transferred via heat exchangers to the main cooling system and then radiated into space by radiators outside the ISS. The flight controller at the systems position has to ensure that one of the two water pumps is running all the time and that there is no leakage in the cooling lines. In the case of anomalies the backup system is activated, or in the case of a total failure most of the subsystems and all experiments are switched off.

6.4.2.4 Data and Communications System

The data of the Columbus subsystems and the payloads are collected by the **data management system** (DMS) on-board the Columbus module. The data is combined in packets and transferred via the US tracking and data relay satellite system (TDRSS) to the NASA ground station at White Sands, New Mexico. The data is then routed to the control center in Houston and relayed to the Col-CC via the ESA interconnection ground subnet (IGS) (see Section 6.4.4 and Figure 6.4.1). The operator on the DMS position is responsible for monitoring and commanding the DMS and monitoring data collection in the Columbus module.

The coordination of data transfer from the Columbus module via NASA and ESA communication lines is the task of the **communication officer**. Due to restricted bandwidth availability, not all parameters of the Columbus module can be sent to the ground in parallel. Hence, the FCT has to select those telemetry packets most suitable for the current task using the relevant predefined parameters. The communication officer is then responsible for switching the packets containing the necessary parameters.

6.4.3 Coordination of ISS Payload Operations

Besides the Columbus system operations, the coordination of payload activities is one of the main tasks

of the international partner (IP) control centers. As already explained, the **Payload Operations and Integration Center** (POIC) at the MSFC in Huntsville performs the payload coordination for NASA. POIC coordinates all payload operations activities in the US segment of the space station as well as all the payload operations for equipment racks integrated in other ISS modules. POIC cooperates closely with the space station control center at JSC in Houston. JSC is responsible for coordinating overall ISS operations and all interfaces (power, cooling, data, video and voice) with the other ISS modules.

Russian experiment operations on the ISS are, as system operations, under the responsibility of the TsUP in Moscow. In contrast to this approach the **European ISS ground infrastructure** has a decentralized structure with the central Col-CC and nine USOCs in different European countries. (see Table 6.4.2 below). The Col-CC coordinates the payload operations in the Columbus module with the user centers and with the control centers of the international partners.

Exploitation of the ISS is a common task of all the international partners. Therefore, close cooperation is required for payload operations and for providing the necessary resources in time. Payload operations can only be started after all relevant tools have been flight qualified and after the formal approval process in all involved control centers has been successfully completed.

6.4.3.1 Payload Coordination at Col-CC

At Col-CC the **operations coordinator** (OC) is the person responsible for experiment operations (Table 6.4.1). The flight controller is in close contact with the flight controllers at POIC and agrees with them the sequence of payload operations. If European experiments are performed in the Russian segment of the ISS, the OC has to coordinate with the flight controllers at TsUP, too. The necessary communication lines for voice, data and video as well as the necessary resources are established and monitored by the systems, the "comms" and the DMS engineers (see Table 6.4.1), together with their colleagues at NASA. NASA is the most important partner because Col-CC has most of its system and payload interfaces with the control centers in Houston and Huntsville.

6.4.3.2 Coordination with the USOCs

The points of contact for European scientists who want to perform experiments on the ISS are the nine USOCs. Table 6.4.2 describes the task distribution among the USOCs. Five USOCs are responsible for all the internal payload racks and external payloads. The others are responsible for single experiments in the payload racks, performing experiments in the US or Russian segments, or supporting other USOCs in their payload operations activities.

The **user centers** support the **scientists** in preparing the experiments for transport to the ISS and during the execution of operations there. Additionally, they provide console positions for scientists monitoring and commanding their experiments or they set up communication lines to the home institutes of the scientists. The commands from the user centers are sent via the ground communication network to the Col-CC and forwarded to the control centers of the international partners and further on to the experiments on-board the ISS. The same interconnections are used for the return link to receive the telemetry

Table 6.4.2: *The European User Support and Operations Centres.*

USOC	Location	Main task
BIOTESC	Zurich, Switzerland	Performance of experiments
B-USOC	Brussels, Belgium	Solar experiments
CADMOS	Toulouse, France	European physiology modules (EPMs)
DAMEC	Odense, Denmark	Performance of experiments
ERASMUS	Noordwijk, the Netherlands	European Drawer Rack (EDR), European Technology Exposure Facility (EUTEF)
E-USOC	Madrid, Spain	Performance of experiments
MARS	Naples, Italy	Fluid Science Laboratory (FSL)
MUSC	Cologne, Germany	Biological Science Laboratory (Biolab)
N-USOC	Trondheim, Norway	Performance of experiments

data from the experiments being conducted on the ISS. The experiment data is recorded and analyzed by the scientists at the user centers or at the relevant universities and research institutes.

6.4.3.3 Coordination with ESA

ESA is one of the international partners of the ISS and is responsible for the European contribution to it. Hence, ESA selects the experiments which are performed on the ISS and places contracts with the USOCs for ISS scientific exploitation. Similarly, contracts are placed with the Col-CC for payload and system operations via the prime contractor Astrium as industrial operator vis-à-vis ESA. When preparing a new ISS mission or a new increment, ESA defines the scope and selects the experiments which are to be performed. ESA negotiates with the international partners on the coordination of all European ISS exploitation activities and places contracts for the transportation of astronauts and payloads to and from the ISS. During operations the FCT is in permanent contact with the ESA operations manager to coordinate and agree on actions in case of unforeseen events or to change operations priorities. Additionally, ESA team members work closely together with the FCT for operations execution: the **ESA MedOps** and the flight surgeons monitor the astronauts' health and set up the necessary training programs; for example, for minimizing the loss of muscle and bone mass of the astronauts under microgravity conditions. ESA also contributes the Eurocom team (see Table 6.4.1), which is responsible for the direct communication with the ESA astronauts on-board the ISS.

6.4.4 The ISS Communication Infrastructure

There are two independent means of communication for routing data, voice and video to and from the ISS. The first one uses NASA's TDRSS, which provides a nearly continuous communication link to the ISS by using one ground station and communicating with the appropriate TDRSS satellites in geostationary orbit (see Figure 6.4.7). The main US communication link to the ISS is routed from the control center in Houston to the ground station in White Sands, and from there to the TDRSS satellites. The signal is then transferred from the satellite to the ISS. Voice, video and data signals from the ISS are sent back the same way.

The second communication link is provided by the Russian communication system, which relies on five ground stations located between the Baltic Sea and the Pacific Ocean. In the optimum case a continuous communication link of more than 20 minutes is possible. The Russian FCT can also use the US satellite communication system when the ISS is out of range of its own ground station network since Russia does not have a dedicated satellite communication system at present.

6.4.4.1 Operations of the European Communication Infrastructure

The **European interconnection ground subnet** (IGS) connects the Col-CC with the control centers of the international partners in Houston, Huntsville and Moscow because the Col-CC has no antennas for communication with the ISS (see also Figure 6.4.7). However, the Col-CC not only routes system and payload data to and from the Columbus module, but also distributes to the ATV control center (ATV-CC), via the IGS, the ATV flight data and the data from docking and undocking activities as well as data generated during docked operations at the ISS. The **automated transfer vehicle** (ATV) is the European resource vehicle (Figure 6.4.8). It is launched by an Ariane 5 from the European launch center at Kourou and can carry up to 10 tonnes of resources (food, oxygen, nitrogen, water and fuel) as well as experiments to the ISS.

Each IGS communication connection consists of a prime and a backup line in order to switch to the backup line in case of failure. The main communication links (data, voice and video) to the ISS and to the Columbus module are routed via the Mission Control Center in Houston. However, the experiment data is provided via the POIC, the NASA payload control center in Huntsville.

The ground communication network has to provide **high reliability** because it is the only communication link for commands as well as voice communication with the astronauts on the ISS. This high reliability of the communication network is

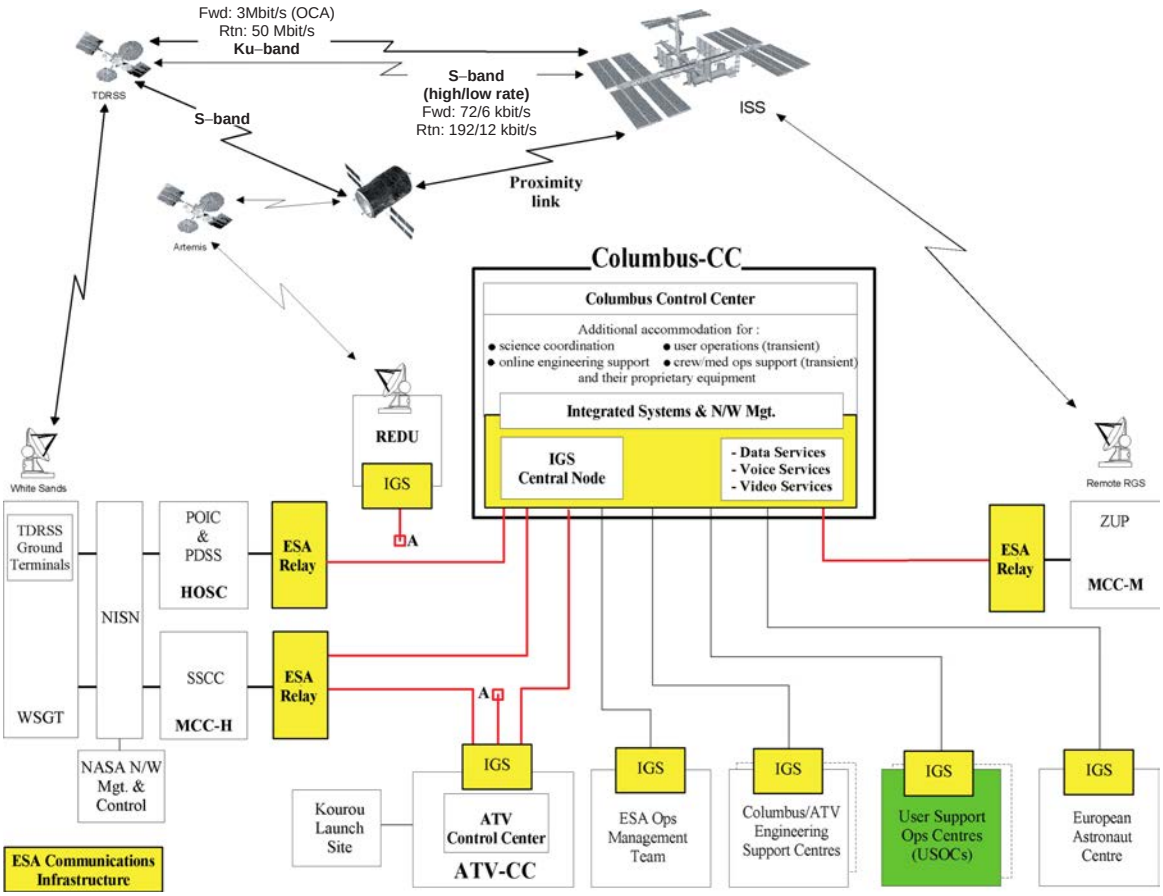


Figure 6.4.7: *ISS communication network (Source: DLR).*

ensured by the Col-CC's ground control team, which configures, operates and maintains the network on a 24-hour basis, 7 days a week. The ground controller is in close contact with the ground controllers at the international partner centers, with the USOCs, and with the control center in Toulouse, which is responsible for the ATV. The ground controller sets up and monitors the communication lines for data, voice and video and informs the FCT about the status of the communication lines. In case of failures the controller initiates measures for recovering the communication network and the ground infrastructure. The ground controller is supported by the system controller, who is mainly responsible for the Col-CC's internal subsystems.

The unmanned ATV (Figure 6.4.8) docks automatically with the Russian segment of the space station. The entire ATV mission is operated and monitored by the ATV-CC in Toulouse. Hence the ATV-CC is connected to the Col-CC by redundant communication lines. The ATV-CC closely coordinates its activities with the Russian control center in Moscow, which necessitates a reliable connection to the Russian control center as well. The communication lines are also used to support the payload operations in the Russian segment and in exceptional cases, for example, if the communication lines to the US control center are not available.

As shown in Figure 6.4.7, for ATV operations the European Artemis satellite (geostationary orbit) is used to augment and secure the ATV data transfer to



Figure 6.4.8: *Automated transfer vehicle (Source: ESA).*

the ATV-CC during critical flight phases. The data is received by the ESA ground station in Redu, Belgium, and routed via a Col-CC-provided interface (IGS node) to the ATV-CC. The IGS node is controlled by the Col-CC ground communications team.

6.4.4.2 Communication with USOCs, EAC and ESCs

The European ground communication network connects the USOCs (Table 6.4.2) as well as the ESCs to the Col-CC. Hence, a scientist located in one of the USOCs can be provided with **data, voice and video connections** to and from the ISS (Figure 6.4.2). After approval by Col-CC the USOCs operate the payloads and monitor the experiments using the telemetry and housekeeping data delivered by the ground communication network. The IGS

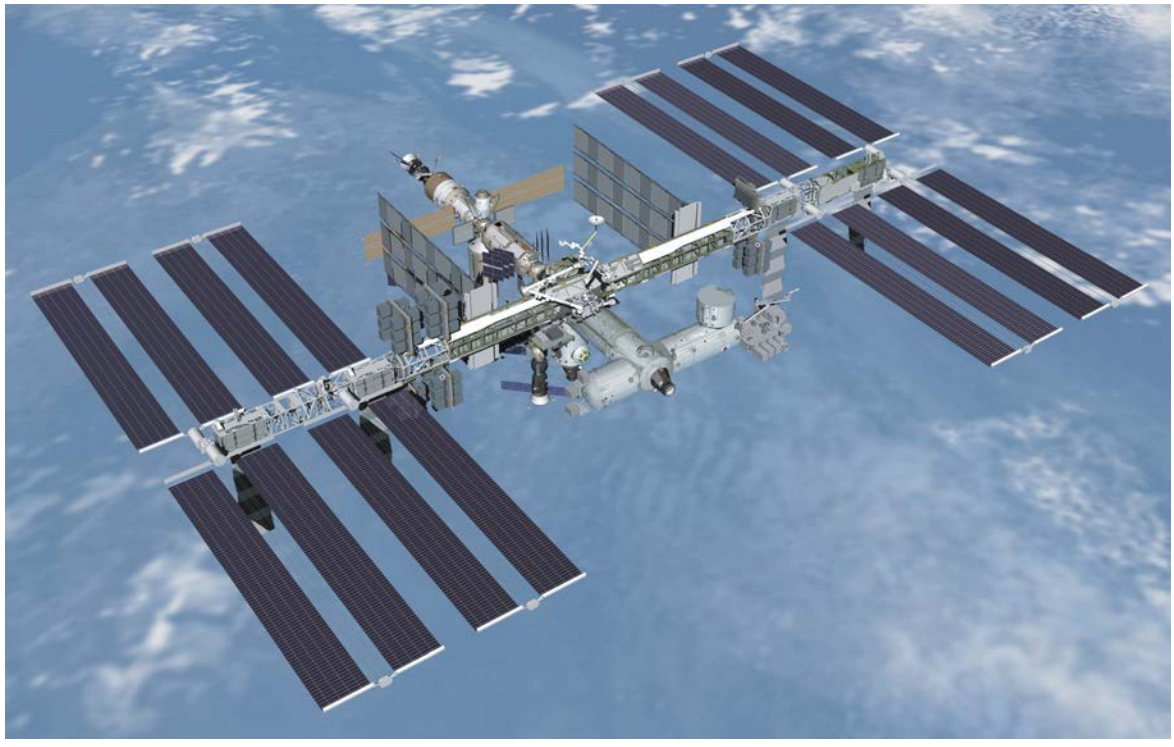


Figure 6.4.9: *ISS after complete assembly (Source: NASA).*

is able to transmit up to 32 Mbps, the bandwidth necessary to transfer the data of a single payload or the combined data rate of several experiments. Some experiments are performed in close cooperation with two independent USOCs which have no direct interconnection. Hence, the communication between

the USOCs is coordinated and routed via the Col-CC. The ground control team in the user centers sets up and monitors the communication lines in close cooperation with the GCT at Col-CC.

The EAC in Cologne is responsible for the **health of the European astronauts** during their stay on the ISS. Data about the health of the astronauts is private and protected, as are the voice channels for teleconferences between the flight surgeon and the astronauts. Hence, access to this data is restricted for use only by the flight surgeons at EAC.

During the launch of the Space Shuttle, the Col-CC routes additional voice channels to the EAC, enabling the flight surgeon to monitor the ascent phase and provide fast medical support in case of an emergency shuttle landing in Europe.



7 Utilization of Space

Berndt Feuerbacher

Even for the early pioneers of space technologies, the utilization of space for **scientific or economic purposes** was a primary goal of their efforts. Nevertheless, **military considerations** were clearly at the forefront during the development of rocket technology. The early phase of space activities in the second half of the twentieth century was marked by technological competition between the two superpowers, the USA and the Soviet Union, and an enormous amount of public financing was accordingly spent on pretentious space projects. But also during this phase, aspects of utilization, particularly those of a scientific nature, were continually emphasized and received the corresponding priority.

Sputnik 1, the first artificial satellite, was launched on October 4, 1957 with a payload consisting only of a transmitter. It had no other function but to demonstrate its existence in orbit – to overwhelming and sensational effect. The first US satellite, **Explorer 1**, launched in January 1958, carried as science payload a Geiger counter designed by James Van Allen. Despite the unexpected saturation of its detector, its measurements led to the discovery of the radiation belts which bear his name, and formed the foundation for the investigation of the Earth's magnetosphere.

Potentials for Exploiting Space

Space offers a wide spectrum of unique qualities and ambient conditions which make it interesting or even essential for a variety of application areas (Figure 7.1). Found here are the new potentials and impulses for

Application area	View toward		Presence in space	Ambient conditions			Cosmic/ global dimensions
	Earth	Space		Vacuum	Radiation	Micro-gravity	
Earth surface	●●●						
Climate and environment	●●●	●					●
Atmosphere and weather	●●●	●					●
Communication	●●●						●
Navigation	●		●				●
Astronomy and astrophysics		●●●	●	●	●●		●●
Planetary research		●●●	●●●	●	●●		●●
Materials sciences				●		●●●	
Life sciences			●●		●	●●●	●●
Technology			●	●	●	●	●

Figure 7.1: *Specific qualities of space and relevant application fields. The number of bullet points represents importance in the respective area.*

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9

scientific and technological research, commercial endeavors, and the basis for necessary activities to meet public responsibilities.

Viewing the Earth from Space

The view from orbit back to Earth has left a deep impression, and not just on the astronauts who were privileged to experience it. The images which showed us our home planet from a new perspective contributed to the change of human paradigms and established a new understanding of our environment. Contiuously improved cameras provide images of the Earth's surface at astonishingly **detailed resolution**, down to 1 m. Presentations such as "Google Earth" are making them easily accessible to anyone and part of everyday life. Less well known are the distinctions now possible with **hyperspectral resolution,** which has vastly expanded information content and is being put into operational use in many areas. The methodology of **radar monitoring**, which provides detailed images of the Earth's surface both day and night, also through clouds, has led to innovative applications in a variety of fields.

Weather maps and forecasts based on satellite data have become a part of the modern news scene. **Weather forecasts** of up to two weeks are inconceivable without meteorological satellites. A look at the Earth's atmosphere from low Earth orbit also provides new insights, since it enables us to monitor our planet's gaseous envelope from its lowest to its highest density, which makes possible measurements with improved resolution. Along with such a global view, **climate effects** can also be detected and reasons for changes identified. One example is the discovery of the ozone hole, which initiated a reversal in anthropogenic influences through a worldwide ban on chlorofluorocarbons.

Technological developments associated with **communication and navigation** in space make use of global visibility from and to the Earth's surface. In both areas commercial applications have gained acceptance and further developments are underway. In the intercontinental voice communication or telephone market, satellite communication was first able to supplant inadequate cable technology, only to be supplanted in turn by fiber optic communication – at least in major population centers. Today there are interesting markets in large developing countries and in mobile communications. Radio and especially television transmission via satellite are economically so interesting that these services are being offered to the user free of charge. This is also the case for available navigation services. Whereas the American GPS system was taken over from the military sector, a European commercial system, Galileo, is being developed with returns expected from value-adding services.

Viewing Space from the Earth

Although Earth's atmosphere appears to be completely **transparent** to the human eye, this is true only for small parts of the electromagnetic spectrum, namely the visible and radio ranges (Figures 7.2 and 7.4.1). From a vantage point beyond the atmosphere, a view of space reveals wide, new spectral ranges which are entirely untransparent or offer only limited access to instruments on the Earth's surface. Space astronomy not only made possible a large number of new discoveries; it also helped to newly define our view of the cosmos.

Electromagnetic cosmic radiation is to be found at the extreme energy-rich end of the spectrum. Imaging technologies cannot be applied here; the information coming from the depths of the Universe is obtained using the measurement methodologies of particle physics. In the **X-ray range**, high-energy cosmic sources such as quasars, pulsars and black holes are imaged and measured spectrally with grazing-reflection optics.



Figure 7.2: *Earth's atmosphere, which appears transparent to the human eye, absorbs most of the electromagnetic spectrum (Source: NASA).*

Hot young stars or the remnants of supernovae are investigated using ultraviolet light.

The long-wave side of the visible spectrum is followed by the **infrared range**. This radiation penetrates the Universe's dust-laden regions and makes it possible to detect cold objects and processes such as the birth of stars. In the **microwave range** Earth's atmosphere is basically transparent. But satellite instruments were nevertheless necessary to measure cosmic microwave background radiation, since it is only the tiniest fluctuations, not measurable through the atmosphere, which provided the decisive information which confirmed our present cosmological model of the big bang.

Presence in Space

Some tasks require a **physical presence** in space. This does not inevitably mean the presence of humans. A robot in space can perform valued services, whether interactively controlled from the ground (**telepresence**) or operating autonomously. Due to safety requirements and the need for life support systems, the technical effort to place humans in space is high. The advantage is in intellectual performance and flexibility, which cannot be achieved by any robot system. But this does not lead to a general preference for either human or unmanned space activities. Depending on the requirements set by the goal of the mission, the most efficient approach has to be selected. It would not make much sense to place a robot in space to conduct medical investigations relating to reversible aging processes, for example.

A presence in space is often required to research **local fields** or **particle fluxes**. Detailed temporal and spatial relationships within Earth's magnetosphere can only be measured by exploiting the simultaneous presence of several probes at various known locations. The first step toward achieving a local presence is to visit objects in our planetary system for the purpose of undertaking instrument-based in-situ studies from nearby. All the planets and several of their moons have been observed during **fly-by** missions or from orbit. The results have considerably expanded our knowledge, and the images which were sent back reveal unexpected, exotic worlds. The next step is to land instruments for **in-situ investigations** (Figure 7.3). The precondition for **landing people** on some objects in space is mastery of return-flight technology. For that purpose there is an interim stage, in which material from a body in our planetary system is brought back to Earth for analysis. This verifies the return technology. At the same time, prior precise investigation of the **extraterrestrial material** increases the safety of humans subsequently placed on the surface of a foreign planetary body. For long-duration presence, resources must be generated locally for fuel, energy and material.

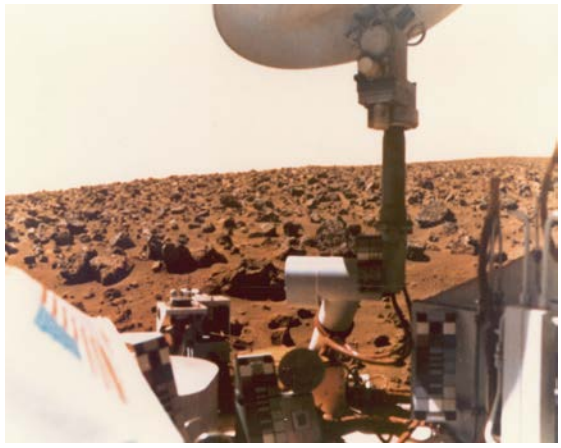


Figure 7.3: *In-situ investigations of planetary bodies contribute considerably to an increase of knowledge. The landing of the Viking spacecraft on Mars in 1976 is a milestone in the research of our planetary system (Source: NASA).*

Ambient Conditions in Space

The **density of the atmosphere** drops exponentially with altitude from its value on the Earth's surface of $3 \cdot 10^{19}$ molecules per cubic centimeter. In low Earth orbits of about 250 km altitude, particle density is still about 10^{10} per cubic centimeter. A decisive criterion for the utilization of space is the question of the stability of the orbit. Some mission goals, such as for Earth observation or communications, call for low orbits. Below 120 km, a satellite's lifetime in orbit, which is also strongly dependent on solar activity, is too short for practical purposes. Above 600 km, atmospheric drag, which depends on the size, shape and orientation of the space vehicle, is no longer the limiting factor for service life, which at this orbit height is over 10 years. Beyond 1000 km, the Sun's radiation pressure exceeds the influences of the remaining atmosphere.

Missions with the goal of utilizing microgravity are likewise limited by the resistance of **atmospheric friction or drag**. The influence of this force depends on the relationship between the mass and atmospheric drag of the space vehicle. Satellites designed for the special purpose of measuring relativity effects or details of the Earth's gravitational field have to be kept in a Keplerian orbit independent of residual atmospheric drag. This is achieved by using sophisticated compensation techniques to follow the position of a free-floating internal test mass.

The ambient **vacuum** is used directly in experiments in plasma physics or molecular physics. For certain scientific experiments, a suitable interface to the outside of a pressurized laboratory module can provide the required vacuum.

The **electromagnetic radiation** in the vicinity of a space vehicle is dominated by the influence of the Sun. The spectrum of the radiation emitted by the Sun ranges from X-rays to radio waves, with a maximum in the visible range at about 500 nm. This is essentially the spectrum of black body radiation at 5800 K, with variations at the two extreme ends of the hard and infrared radiation ranges. At these borders one finds distinct fluctuations caused by solar activity, whereas the total flux, expressed by the solar constant of 1371 W/m^2, remains basically unchanged.

Outside the protective envelope of the atmosphere, a space vehicle is exposed to the total **particle flux** of space. This consists of elementary particles, ions and atomic nuclei. In the high-energy range this particle radiation penetrates organic as well as inorganic matter and can cause damage which affects the health of astronauts or causes equipment to malfunction.

This particle flux comes almost entirely from the Sun in the form of the solar wind, which is composed of protons, helium nuclei (alpha particles) and electrons. With flux velocities of 400 km/s this is equivalent to kinetic temperatures around 10 000 K or energies of 1 keV. Solar storms or eruptions can release significantly higher amounts of energy, which can approach the realm of cosmic radiation at over 10 GeV. With increasing energy, the particle flux of 10^{15} per square centimeter and second at 1 keV drops by 18 orders of magnitude in the gigaelectronvolt range.

Gravity is ubiquitous in space. It originates from the overlapping of the gravity fields of all the masses distributed throughout the Universe. It can be neither screened nor switched off. Since this force decreases with the square of distance, our planetary system is dominated by the gravity of the Sun, which has 750 times as much mass as all the other planetary bodies taken together. Only in the immediate vicinity of another body does that body's own gravity field predominate.

Earth's gravity field at the altitude of stable orbits, about 250 km, is still about 93% of the value on the Earth's surface. Weightlessness arises through compensation of the remaining gravity influences by the inertial forces arising from the satellite's motion, which in the case of an orbit coincides with the centrifugal force.

Microgravity in space was an attractive phenomenon even in the early days of space activity. On the ground, equivalent conditions can only be achieved with drop towers, parabolic aircraft flights and suborbital rocket flights, but only with inevitable restrictions as to the time available for experiments and the quality of the gravity reduction. Under conditions of microgravity, the effects of sedimentation, convection and hydrostatic pressure disappear, which leads to novel conditions, particularly for studying processes involving at least one fluid component. In the research fields of materials sciences, medicine and biotechnology, as well as in the field of basic physics, innovative results can be obtained and new approaches developed for industrial processes on the ground.

The European space laboratory **Spacelab** (Figure 7.4) was designed primarily for scientific research under conditions of orbital microgravity. With its 25 missions it was the Space Shuttle's most frequently flown payload and was the foundation for the development and utilization of the International Space Station.

Global and Cosmic Dimensions

In space there are no limits placed on the fantasy and inspiration of scientists and engineers. In the vastness of space, coupled with an absence of gravity effects, structures of gigantic dimensions can be imagined. Wernher von Braun had already developed in the 1950s concepts for huge **space stations**; even artificial colonies to house fractions of Earth's population were envisioned.



Figure 7.4: *Spacelab, Europe's space laboratory, was put to use during 25 space missions. Here astronaut Ulrich Walter demonstrates microgravity during the D2 mission of 1993 (Source: DLR).*

The global dimension of space activities is a generally accepted, prevalent concept. This is evident in the fields of **Earth observation and climate research**, which are used to diagnose environmental problems as well as contribute to the development of suitable countermeasures.

Science projects exploit the vastness of space in countless ways. Observations using **very long baseline interferometry** (VLBI) make possible spatial resolutions of the Universe which cannot be represented in any other way. The search for planets outside our Solar System of a size similar to Earth's will make good progress thanks to such methodologies. A concrete joint NASA and ESA project planned for launch in 2015 is a gigantic laser interferometer (LISA, Laser Interferometer Space Antenna). Comprising three spacecraft flying 5 million kilometers apart and together forming an equilateral triangle, LISA is intended to open up a new window for observing our Universe by measuring gravity waves.

7.1 Earth Observation

Klaus Dieter Reiniger and Gunter Schreier

Earth observation is an important space application, and during the pioneering years of space flight it was a primary goal of space utilization. Early attempts by the USA to conduct military reconnaissance using camera-equipped satellites flying over the Soviet Union are known. Improvement of weather forecasts based on observations of weather phenomena from space is now a matter of course, and even high-resolution images of the Earth on the Internet are now drawn on by a large number of the most varied users.

7.1.1 Categories of Earth Observation Applications

Weather Observation and Atmospheric Conditions
Monitoring the momentary **state of the weather** from space is one of the oldest applications for satellite technology. The USA's Vanguard 2 satellite (launched in 1959) measured the reflection of sunlight from the Earth and clouds (albedo) as part of the International Geophysical Year (IGY) and can be regarded as one of the first systems for monitoring weather and climate parameters. In 1960 the US TIROS satellite provided the first black-and-white images using a video camera. The successor system already included infrared spectral range observations. The sequence of weather and environmental satellites in low Earth polar orbits continued in 1970 with the satellite series of the US weather agency NOAA. With its **advanced very high-resolution radiometer** (AVHRR, with approx. 1 km geometric resolution on the ground) the **NOAA series** of weather and environmental satellites still provides the basis for mapping many atmospheric, land and water parameters (e.g., vegetation indices, ground and water surface temperatures) on a daily, global basis (see Figure 7.1.1). After 2006 the MetOp satellite of the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) complemented this fleet and added additional functionality, particularly for measuring atmospheric **trace gases**.

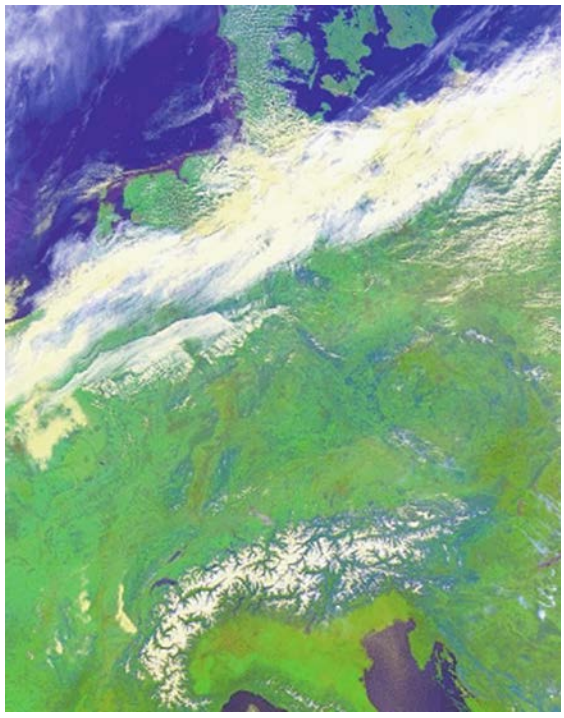


Figure 7.1.1: *NOAA-17 image of Central Europe, recorded on March 16, 2007 at 10:19 MEZ (Source: NOAA/DLR).*

Trace gases such as ozone, NO_x, SO_x and others can be globally measured from polar orbiting satellites primarily using spectrometers (GOME, SCIAMACHY on the ERS satellite, ENVISAT and MetOp). Especially, measurements of "greenhouse gases" over many years reveal important parameters for analyzing and predicting global climate change.

Almost constant weather monitoring for given locations on Earth from space is possible with **geostationary weather satellites**. Since the launch of the US GOES-A satellite in 1975 further geostationary weather satellites have been added. At present (2008) the USA/NOAA (135°W and 75°W), EUMETSAT (0° and 63°E), the Russian Federation (76°), China (105°) and Japan (140°) operate geostationary weather satellites, whose data provide a view of the entire hemisphere at up to 15 min intervals with a resolution of 1 km in nadir (this information relates, for example, to the SEVIRI sensor on the newest Meteosat satellites).

The data from geostationary satellites is usually received at central ground stations, processed, and in some cases transmitted in processed form directly via a transponder on-board the same satellite. The EUMETCAST service of EUMETSAT allows digital reception of additional environmental parameters with an 85 cm Ku-band antenna and an appropriate PC interface.

Besides the climate and weather data generated by national weather services and increasingly also by private suppliers, all data is combined into a total picture at international data centers. The **European Centre for Medium-Range Weather Forecasts** (ECMWF) in Reading, UK, is one of the institutions which generates long-term weather forecasts from satellite data and models of the atmosphere.

Monitoring the Environment

Monitoring changes on the **land surface**, especially **vegetation** related to agriculture and forestry, has been a focus of interest in the utilization of remote sensing data since the 1970s. Up until that time, weather satellites supplied daily images of entire continents at low geometric resolution. With the launch of the **Earth Resources Technology Satellite** (later renamed **Landsat 1**) in 1972 the era of environmental and resource mapping began (Figure 7.1.2). The multispectral scanner (MSS) on the Landsat series of satellites provides ground resolution of 80 m in four spectral channels. Ten years later, on Landsat 4, it was complemented by the thematic mapper (TM) with a ground resolution of 30 m in six spectral channels (plus a thermal channel with 120 m resolution).

Russian systems, and later especially the French **SPOT program**, presented more possibilities to monitor large parts of the Earth's surface from space at higher geometric and multispectral resolutions. At the end of the 1990s, Indian, Chinese and increasingly commercial satellite systems were added to this observation fleet. Of the last, the German **RapidEye system** with its five satellites should be mentioned; with its 6.5 m geometric resolution it will set new standards (launched in 2008).

The applications for data coming from such systems, starting with the Landsat satellites, include estimates of **biomass**, both natural vegetation and agricultural crops. Other applications make use of remote sensing data on plant biodiversity, bare ground formations, rock types and human settlements. Land use change

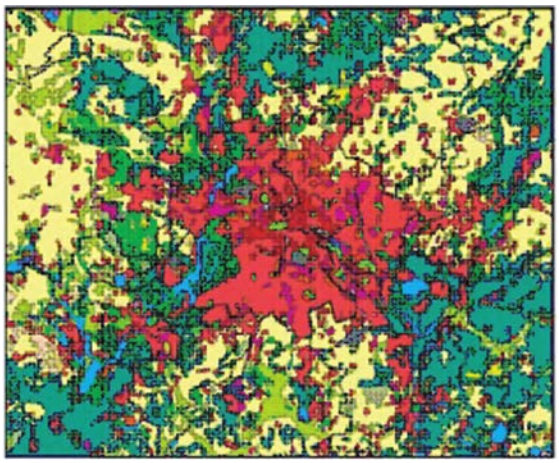


Figure 7.1.2: *CORINE 2000 classification of Berlin and surroundings based on Landsat data. The color coding indicates different land use classes (Source: Eurimage).*

detection studies provide one of the foundations for demonstrating global change.

Since this data has been available from the early 1970s – in most cases at low cost or even free of charge – it has become the cornerstone of many national and international **environmental mapping programs**. Thus the European Commission uses them in its CORINE program (Coordinated Information on the European Environment program) to regularly record land use in Europe, currently on a five-year basis.

Agencies of the United Nations use the data to monitor specific measures related to development aid and global supervision of human settlement patterns. Deforestation as well as reforestation is at the focus of discussions on how to manage the global carbon dioxide budget.

Calls to map the land surface have increased since this technology has become available. High geometric resolution is necessary for detailed recording of small-scale changes in agricultural areas or the growth of cities. For even more precise recording of vegetation types and their condition, as well as geological structures, better **spectral resolution** is called for. In addition to multispectral measuring equipment (with up to 12 channels), hyperspectral sensors with over 200 spectral channels will be required in the future. The German EnMAP satellite will meet this need starting in 2012.

Monitoring Geophysical Parameters

When approaching Earth from outer space, one of the first geophysical parameters one encounters and can measure when still at a distance of hundreds of thousands of kilometers is its magnetic field. But Earth observation satellites move in much lower orbits, where they are subject to a variety of other physical influences, including complex geomagnetic fields and radiation. Nevertheless, from this vantage point and with simple radiation measuring equipment, the very first US satellite (Explorer 1, 1958) discovered the Van Allen radiation belt.

Since that time, magnetometers to measure **Earth's magnetic field** have flown on a number of satellites. At present (2009), three satellites, including Germany's CHAMP, carry magnetometers. With their orbit geometry especially attuned to this task, the ESA SWARM mission should be able to measure the inner and outer components of the Earth's magnetic field starting from 2010.

Gravity is a critical field for global and regional geophysical phenomena, including global climate indicators. Masses and mass transport in the Earth's core and on its surface deform the so-called geoid, a mathematically determined surface which represents areas of equal gravitational force. One of the largest anomalies is to be found in southern India, where the discrepancy from a regular ellipsoid is about 110 m (see Figure 7.1.3).

Since Earth-orbiting satellites orient themselves according to the local gravity field, even small variations cause minimal changes in the satellite's orbit. Very precise measurement of the satellite orbit (orbitography) is accordingly the key to calculating the gravity field. Measurement methods include, for example, **laser tracking**, in which reflectors on the satellite reflect back laser pulses sent from the Earth's surface. A global network of satellite laser tracking stations (SLRs) carries out these measurements. But the position and velocity of a satellite are also affected by numerous other influences which must be taken into account – not least the remaining atmosphere, even at high altitudes. Besides laser tracking, microwave signals and GPS positioning are used for precise orbit determination.

Conversely, precise knowledge of the gravity field is used to **calculate satellite orbits**. Exact attitude

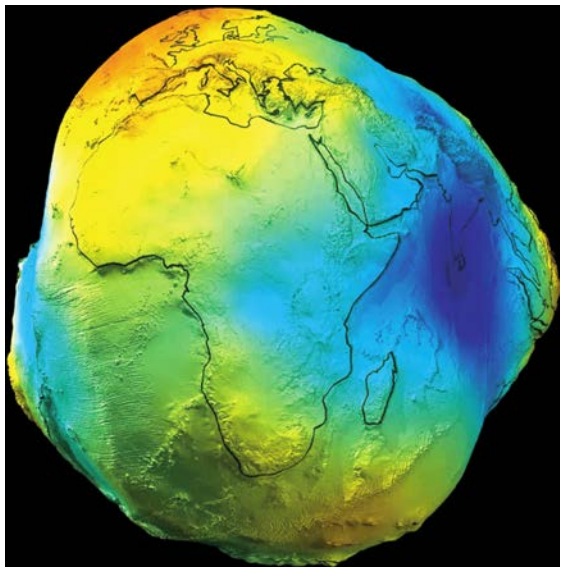


Figure 7.1.3: *Height-exaggerated visualization of the geoid based on measurements from the CHAMP satellite (Source: GFZ Potsdam).*

determination and thus the precise location of ground pixels in the case of optical satellites is possible thanks to gravity field measurements. But also changes in climate, such as the El Niño effect in the Pacific Ocean, can be tracked down as differential changes in the gravity field. Thus evidence for their existence comes from the methodology of satellite geodesy.

Mapping the Land Surface

The conversion of images of the Earth's surface into accurate cartographic representations is the subject of **photogrammetry** (also known as aerial photography surveying). Driven by government planning and military requirements and supported by the evolution of aircraft and improved camera technology and film material, photogrammetry is a child of the early twentieth century. At that time, analog images were turned into maps using a combination of optical and mechanical approaches. Stereo-optical images were used to calculate the heights of topographic features and other objects.

The first – including military – **Earth surveying systems** from space also depended on analog cameras and subsequent photogrammetric interpretation (e.g., the metric camera on the 1983 shuttle mission). The first civilian electronic systems had geometric resolutions of 80 m (Landsat 1, 1972) and 30 m (Landsat 4, 1982) and thus lacked the geometric precision required for generating maps at smaller scales. This meant that structures such as streets could not be recognized. Also, these instruments did not allow stereo-optical oblique views of the Earth. This situation changed with the launch of the French SPOT 1 satellite in 1986. **Geometric resolution** of 10 m coupled with the ability to point the sensor's field of view made it possible for the first time to generate **three-dimensional topographic maps** at small scales. In addition, analog photogrammetric methods could be replaced by new digital algorithms and procedures. With the fall of the Iron Curtain geometric resolution better than 10 m was possible for civilian applications. Besides the first Russian data, the launch of the civilian US satellite IKONOS-2 in 1999 marked the beginning of a new era of high-resolution optical satellites with resolutions better than 1 m (see also Table 7.1.2 below).

The **sensors** on these satellites are optimized for the purpose of rapid mapping. High geometric resolution is in most cases achieved only in the panchromatic channel (black-and-white images). A few other channels with lower geometric resolution permit false-color representations. When the data is used as the basis for maps, the **precision of the attitude control** of the sensor or satellite is important. Complex control systems usually permit both rapid and precise orientation of the entire satellite. Also, numerous orbit- and attitude-regulating sensors assure the precise location of a pixel on the Earth, down to a few meters. Further improvements permitting extension down to the meter range and below can only be achieved by making a connection with known terrestrial targets ("ground control points") such as street crossings. Rapid modification of a satellite's orientation also makes it possible to record a given scene once again from another oblique angle and thus to have a stereo pair for three-dimensional analysis. If two cameras with different orientations are available on-board one satellite, such stereo-optical images can be regularly provided.

Besides generating and updating topographic maps, the spectral information provided by these satellites makes detailed **thematic mapping** possible. In the civilian realm, basic information for planning

major constructions can be provided or new streets identified for updating car navigation systems. High-resolution mapping of crops makes it possible for the European Commission to confirm the appropriateness of agricultural subsidies, or the United Nations to locate illegal drug cultivation. Certain nuclear activities worldwide can be revealed with high-resolution images.

Rapid mapping based on high-resolution satellite data increasingly supports humanitarian aid activities in crisis areas and after natural catastrophes (Figure 7.1.4). When compared to older data, current data reveals the destruction of infrastructure after an earthquake or the extent of flooding. Because of the urgency of such information and the need to be independent of clouds and time of day, **high geometric resolution radar systems** (e.g., TerraSAR-X, COSMO-SkyMed) are increasingly being used.

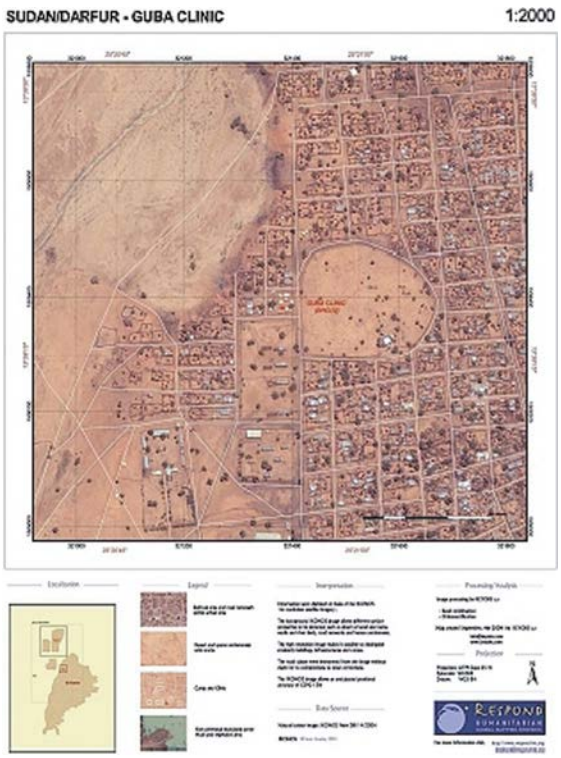


Figure 7.1.4: *Satellite image map (1:2000) showing the Al Fashir/ Abu Shok refugee camp and used for humanitarian aid in Sudan in 2004 (Source: DLR-DFD, ZKI).*

Monitoring the Oceans

Over two-thirds of the Earth's surface is covered by oceans. They are among the most important factors in day-to-day weather and long-term climate development. Direct measurement of parameters for the state of the ocean surface can only be provided for specific locations because of the enormous distances and depths involved (anchoring of buoys). Therefore, Earth observation satellites can provide important information on these parameters. Satellite-based measurement methodologies for ocean parameters cover almost the entire range of applications currently possible with satellites.

In the **optical range**, narrow-band radiometers and spectrometers are used to map the ocean surface in several spectral channels. Pure water is only "visible" in the wavelength range between 400 and 750 nm. In this range, suspended matter (sediments), chlorophyll (plankton algae) and other organic and inorganic material can be recorded (Figure 7.1.5). In the **thermal range**, sea surface temperatures (SSTs) can be measured. When mapping oceans, global daily coverage is usually desired. Optical sensors thus have broad-swath widths and low geometric resolutions (1150 km at 300 or 1200 m resolution in 15 spectral channels for MERIS on ENVISAT, for example).

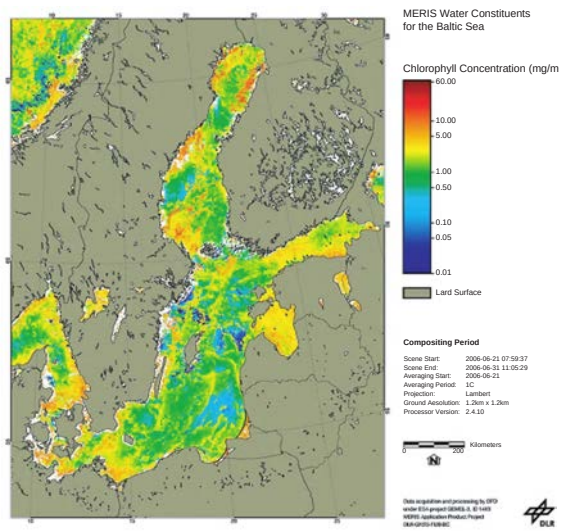


Figure 7.1.5: *Map showing chlorophyll concentration in the Baltic Sea based on data from the MERIS sensor on the ENVISAT satellite (Source: ESA/DLR).*

Radar sensors used as altimeters measure **ocean topography**. The US–French mission TOPEX/POSEIDON can detect 4–5 cm height differences, for example. Ocean currents ("the Gulf Stream") and surface winds can also be measured with this technology.

Active **imaging radar sensors** (**synthetic aperture radar** (SAR)) have been used since 1978 (US Seasat) to map the ocean surface and polar ice. Since water acts like a mirror for microwaves, regular wave patterns can be traced in SAR images. These patterns enable conclusions to be drawn about wave height and direction and the velocity of surface winds. Such analyses have applications in the optimal sighting of off-shore wind farms or in detecting oil slicks (because of the attenuation of surface waves). Special reception modes ("wind and wave mode") on the European satellites ERS and ENVISAT can be employed to generate small samples ("imagettes") along the satellite track; taken together these permit regular ocean monitoring.

Security and Military Applications

Most of the systems, technologies and methodologies of satellite cartography come from military programs. The first photogrammetric cameras on satellites flew on top-secret missions of the USA and former USSR. On August 18, 1960 the first **military reconnaissance mission** of the USA, the Discoverer 14 satellite, was launched. The exposed **analog film** was ejected in a capsule which was captured in flight by dedicated aircraft. The geometric resolution was 5–10 m. Because of the typically low orbits and the high consumption of film material, these missions had a short life. The entire program, named CORONA, flew about 95 missions up to 1972. The concept of analog film ejected in film capsules was replaced in the mid-1970s by electronic systems. Meanwhile, the 800 000 analog images have been digitized and made available to the public.

The former Soviet Union also operated reconnaissance satellites. ENIT-2, launched in 1962, even had separate lenses to be able to record in six different spectral channels. The camera, built by Carl Zeiss in Jena, recorded these channels on different films of 70 mm width and 120 m length which were likewise ejected in capsules.

Naturally enough, there is no detailed information available about current military reconnaissance systems. Nevertheless, the latest generation of US reconnaissance satellites in the optical range is said to achieve **geometric resolution below 10 cm**. Using geostationary data relay satellites these images are available in real time and presumably also as video sequences. The technical design and size of these satellites can be compared to the Hubble Space Telescope. With enough fuel on board, they can also change their inclination and orbit height during their service life, in order to reach given targets on Earth earlier or more frequently. The satellites operated by the US National Reconnaissance Office (NRO) go by names such as **Keyhole**.

Because optical systems do not have **all-weather capabilities**, security applications instead require active radar systems (SAR), which can provide images both at day and night in all sorts of weather. The US military have a few such systems ("Lacrosse") in operation. The possibility of also using SAR to detect **moving objects** gave rise to a classified US project to launch a fleet of SAR satellites into orbit in order to record in real time all moving objects on the Earth. At present its realization is precluded by the enormous costs involved and by the very large amounts of data which would have to be managed by the ground segment.

As an effect of its development of the SPOT series of Earth observation satellites, France, as the only country in Europe, operates optical reconnaissance satellites which are only available to the military (Helios 1 and 2). However, cooperation agreements make the data also available to other countries, including Germany, which launched a military SAR system (SAR-Lupe) at the end of 2006. By contrast, a "dual-use" system is represented by the Italian COSMO-SkyMed SAR satellite, which was launched in 2007. The European fleet of Earth observation satellites is complemented by the French PLEIADES system with a geometric resolution of some 70 cm. It is available for military as well as civilian use.

In addition, both military and security authorities also use data from civilian satellites to meet their information requirements. For that reason the US commercial systems IKONOS and QUICKBIRD as well as their successors GEOEYE and WORLDVIEW are primarily financed by customers from the military realm.

7.1.2 Elements of Earth Observation Missions

The requirements placed on Earth observation missions come from users as well as systems considerations. A primary task during the definition and design phase of Earth observation missions is to balance both sets of requirements and develop a coordinated mission concept.

For the user, these requirements relate in the first instance to subsequent applications for the mission's sensor data [7.1.2], whereas the concern of the systems side is technical realization of these requirements, the satellite and its subsystems, the ground segment and later on mission operation. The requirements can be organized into four main groups.

7.1.2.1 Observation Profiles

Often the user's interest in Earth observation data is in having access to **observation areas of interest** which may be scattered over the Earth's entire surface. The user's application usually focuses on monitoring large areas: for example, following the transformation into steppes or growing desertification, monitoring the annual vegetation cycle, or keeping track of changes in the polar ice caps. But the user's interest may also frequently be in having very detailed information about selected small areas of the type to be found in urban settings, or in discrete objects such as airports, or even individual buildings. These requirements directly influence the mission profile, the design of the orbit and sensing system, and the subsequent data processing. Corresponding to this **observation profile**, there are different requirements for the orbit, the sensing instrument and the operations plan.

7.1.2.2 Typical Orbits of Earth Observation Satellites

In general, geostationary as well as circular or even elliptical near-Earth orbits are suitable for Earth observation missions. However, the latter are typically employed for very specialized observation profiles such as reconnaissance missions.

Geostationary Orbits

The geostationary orbit is used primarily for permanent and wide-area **observation of the Earth's surface** and its **atmosphere** for applications in **meteorology**, **climatology** and **environmental investigations**. Satellites in this orbit are able to permanently monitor a large part of the Earth. Three satellites equally spaced over the equator assure a permanent observation constellation for more than 90% of the Earth's surface. However, the polar regions cannot be included. With the satellites' continuous observation capability and their fixed position with respect to the Earth's surface, dynamic processes in the atmosphere such as cloud formation, strong wind phenomena such as cyclones, and the movement of large masses such as desert sand can be well detected and monitored. Because of the satellites' great distance from the Earth's surface, the **spatial resolution** of optical sensors is **limited** to about 1 km, and relatively long sensor integration times are necessary. Typical representatives of this satellite type are geostationary weather satellites such as Meteosat and GOES.

Low Earth Circular Orbits

This orbit is primarily used for applications with **medium to very high geometric resolution** of objects on the Earth's surface. These satellites use a corridor of about 430–930 km altitude and orbit the Earth with a period of 90–110 minutes, in other words between 13 and 15 times a day. The monitoring of the Earth's surface is therefore not continuous as in the case of geostationary satellites, but rather occurs in daily flights over the target area. The number of overpasses and thus the **daily observation duration** as well as the **contact time** with a ground station are limited and depend on the geographical position. Contact times are at most 15 minutes during a direct overpass. The preferred orientation for circular orbits is over the poles, because the Earth's rotation under the satellite orbit plane makes it possible to sense the Earth's entire surface at regular intervals. An important special case of this orbit orientation is the so-called **Sun-synchronous orbit**. With careful selection of altitude and inclination (approx. 98°), the precession of this orbit is equal to one rotation per year. Thus the orientation of the orbit is always the same with respect to the Sun, which means that a location on the Earth's

surface is flown over at the same local time each day, so the solar observation angle remains constant. The result for optical satellites is comparable conditions of illumination for daily observation at the target location. Some radar satellites also use Sun-synchronous orbits by flying along the border between day and night (**dusk–dawn orbits**) and do not pass over the Earth's shadow. This makes it possible for their solar generators to be continuously illuminated by the Sun, which is beneficial for the charging of the batteries.

However, the required high spatial resolution can only be achieved by limiting the width of the sensor scan, the so-called **swath width** around the footprint on the ground. Typical swath widths are 20–85 km for land surface observation systems (SPOT, IRS, Landsat) and 10–21 km for systems with object-detection capabilities (IKONOS, QUICKBIRD). By lining up scanned swaths next to each other, one can obtain a complete image of the Earth's surface within one observation cycle. As to the orbit path, for almost complete daily coverage a so-called **floating orbit** is necessary, which leads to continuous offsetting of subsequent footprints on ensuing overpasses. By contrast, a **frozen orbit** produces a daily integral repeat pattern. The time interval required for complete coverage of the Earth's surface is called the **repeat cycle** and describes the time that passes before a particular ground track is repeated. It should be distinguished from the **revisit time**, which describes the time before some location on Earth can be sensed again. Since this time is a function of the geographic latitude of the target area and the capabilities of modern satellite systems to rotate up to 60° from their nadir track, the revisit time is always lower than the repeat cycle.

7.1.2.3 Sensor Systems

For Earth observation, primarily **optical sensors** and passive and active **microwave sensors** are used. The goal of the recording system is to achieve as far as possible an image of the Earth with moderate to very high spatial resolution. For both sensor systems the currently preferred resolution range for civilian Earth observation satellites is between 5 and 30 m for large-area monitoring (Landsat, SPOT, ERS, ENVISAT, RADARSAT-1), as well as the range between 0.5 and 3 m for detailed imaging (QUICKBIRD, IKONOS, TerraSAR-X, RADARSAT-2). Typically, both types of sensors offer different sensing **modes**. These can produce either a continuous, "strip map" or a selective, scene-oriented recording. For the latter, a "look and stare" or "spotlight" mode is frequently used, in which the sensor limits its scan onto the target area by mechanical or electronic means in order to obtain an acceptable signal-to-noise ratio.

Both optical and microwave sensors provide technical means for measuring the topography of the Earth. Whereas optical sensors follow the classic **stereo method** of using fore- and aft-looking cameras or different satellite passes for sensing from different observation positions, radar sensors primarily use **interferometric methods** with two receivers at different locations or with two overpasses of the same receiver in a multipass procedure.

Optical Sensor Systems

The classic optical multispectral recording processes use either mechanical scanning (Landsat) or electronic **line scanners** with one to five rows (MOMS). Multispectral scanners usually employ 4–7 spectral bands in the visual and thermal wavelength range (Landsat 7) selected according to the desired application (Table 7.1.1).

Frequently, a **panchromatic sensor** (0.45–0.90 µm) with significantly higher geometric resolution is added to sharpen the multispectral image (IKONOS, SPOT). For modern optical systems the radiometric resolution is 10–12 bits and permits object recognition even in shaded areas. **Optical cameras** are today primarily equipped with optical telescopes with focal lengths

Table 7.1.1: *Spectral ranges of the Landsat TM sensor.*

Color	Wavelength [µm]	Application
Blue	0.45–0.52	Water, forest
Green	0.52–0.60	Intact vegetation
Red	0.63–0.69	Plant species, geology
NIR	0.76–0.90	Biomass
SWIR	1.55–1.74	Water content of vegetation
TIR	10.4–12.5	Vegetation stress
MIR	2.08–2.35	Plant and soil moisture

of several meters and apertures of 40 cm to 1.2 m. As imaging sensors, linear CCD rows with over 12 000 pixels are used.

Optical sensing systems (Figure 7.1.6) with very high, so-called **hyperspectral resolution** of 200 or more spectral bands have so far only been employed to monitor the atmosphere (GOME, TOMS, SCIAMACHY). As they are used primarily to measure the concentration of trace gases in the atmosphere, their moderate geometric resolution (40–200 km) is adequate. The value of this type of sensor for land surface or ocean applications is, however, evident. Progress in technical developments will soon support their application in spaceborne systems (EnMAP).

Infrared sensors are being integrated in multispectral and hyperspectral instruments. The spectral ranges of reflective (3.7–5.4 μm) and emissive (12.3–15.4 μm) infrared are of particular interest. The TM on Landsat 4/5 was equipped with spectral channels in these infrared ranges and the planned sensor suite for the hyperspectral EnMAP satellite will likewise be equipped with several channels in the infrared range. Another special application of infrared sensors is recognizing **hotspots** like forest and bush fires, volcanic eruptions and disasters at technical facilities (Chernobyl). DLR's small satellite BIRD was in the past few years able to successfully demonstrate the value of new infrared sensor technology for applications-oriented missions. Further fire-detection missions based on this satellite are being planned.

Microwave Sensors

In contrast to optical systems, microwave sensors do not require solar illumination of the target and can take **measurements through clouds**. They are accordingly preferred for monitoring areas of the Earth which are cloud covered (as in the tropics) or do not receive sunlight (polar regions) for a high proportion of the year. Because of the need to send out their own illumination signal to the target area, active radar sensors using an SAR approach are technically more elaborate than optical sensors, and are thus usually limited to one radar frequency range. This disadvantage is partially compensated by the **multiple polarization capability** which is increasingly being used in new sensors, as well as in **formation flights** of satellites bearing radar sensors of different frequency ranges. Such formations were originally planned for the TerraSAR-X (Figure 7.1.7) and TanDEM-X missions and will presumably be realized with the Italian X-band mission COSMO-SkyMed and the Argentine L-band mission SAOCOM. Because of the different absorptive characteristics of the microwave radiation, the **frequency ranges** of L-band (2 GHz; TerraSAR-L, SAOCOM) for vegetation monitoring applications, C-band (5 GHz; ERS, RADARSAT, ENVISAT) for ice and ocean monitoring, and X-band (9 GHz; TerraSAR-X, COSMO) for purposes of high-resolution object detection are preferred. Radar sensors usually employ active **array antennas** with a large number (384 modules in the case of TerraSAR-X) of transmit and receive modules as emitters, or concentrating reflector dishes.

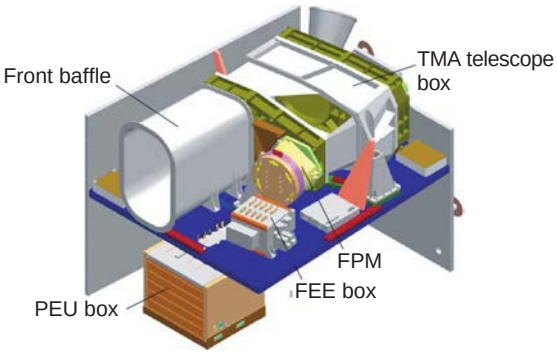


Figure 7.1.6: *Optical camera JSS56 of DJO for RapidEye (Source: RapidEye).*

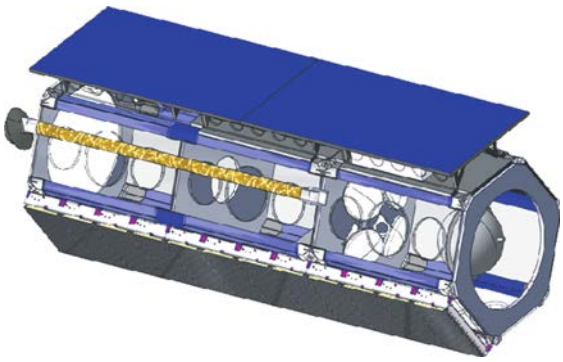


Figure 7.1.7: *TerraSAR-X satellite in flight configuration (Source: ASTRIUM).*

In addition to these sensors, **passive microwave radiometers** are used to determine the water vapor content of the atmosphere [7.1.4], polar ice cover [7.1.5] and precipitation in the tropics [7.1.6].

In the future, lidar sensors will also be used for Earth observation from space, especially to measure wind fields. For this purpose, ESA has approved the ADM-AEOLUS mission [7.1.7].

7.1.2.4 On-board Data Handling

Modern Earth observation sensors generate data at rates and volumes which place extremely high demands on handling and transmission of the data to ground. **Peak data rates** of several gigabits per second are no exception for modern Earth observation sensors and cannot as such be transmitted directly to ground, since the required transmission bandwidth is not available due to international regulations. **Compression** of the raw sensor data and interim storage in so-called **solid-state mass memories** (SSMMs) on-board the satellite are therefore essential. Frequently, data processing also includes **encryption** for the purpose of security from unauthorized access. The **transmission** of data from satellite to ground takes place in the dedicated transmission band (X-band, 8.025–8.4 GHz) at the current maximum data rate of 320 Mbit/s. The available **bandwidth** is sufficient for data transmission from most Earth observation satellites today. Future systems will, however, require new transmission technologies, such as bandwidth-optimized modulation schemes, multiple polarizations, a switchover to new transmission bands, and the transfer of data via geostationary communications satellites at very high-frequency transmission (Ka-)bands or via optical transmission links (e.g. Laser Communication Terminal).

7.1.2.5 Payload Ground Segment

The logical connection between the sensor system on-board Earth observation satellites and the users of the sensor data on the ground is the so-called payload ground segment. It comprises all the functions and services necessary to supply the users with the information they desire and to make it possible for them to put it to the intended use. These functions include order placement by the user as well as mission planning and sensor tasking, data reception, processing, archiving, and distribution of information to the user.

The **technical structure** of a payload ground segment is quite different from that of the flight or mission operation segment, which is responsible for the satellite platform, mission planning and sensor operation. There are, however, numerous **interfaces** between both facilities at the level of mission planning and operation. The technical elements have much in common, as do the terrestrial communications, ground stations, databases and archives; the main differences are in the techniques used to handle the **data rates and volumes** from the sensor. These are frequently several orders of magnitude larger than is the case for the mission operations. Transmission rates of several hundred megabits per second are today the state of the art and will grow in future by another factor of 10. This is the reason for differences in the technical design of the payload ground segment, which is oriented toward the provision of user services, data sets and information products. These functions and services are making **data access in near real time** increasingly possible and have the goal of generating input products for geophysical and Earth-related analyses. Another consideration is that in many cases new physical insights can only arise from completely new **reprocessing of data sets** which incorporate long time intervals of monitoring. The most obvious example is the discovery of the ozone hole over the south polar atmosphere, which, although detected in the first data from the CCSZ sensor on-board the Nimbus satellites, was only recognized after complete reprocessing of the data with new algorithms.

The establishment of a payload ground segment is a significant step in realizing an **Earth observation capacity**, since timely data availability and long-time continuity of data archiving are primary preconditions for exploiting satellite-supported Earth observation activities. This is true for applications of immediate interest, like provision of information during catastrophes, as well as for evaluating data revealing long-term changes in vegetation or urban landscapes, for example. Thus the preservation and maintenance of **large data archives** from various data sources is a central task of payload ground segments, a task which has been jointly carried out in Europe for a number of years by several data centers in collaboration with ESA.

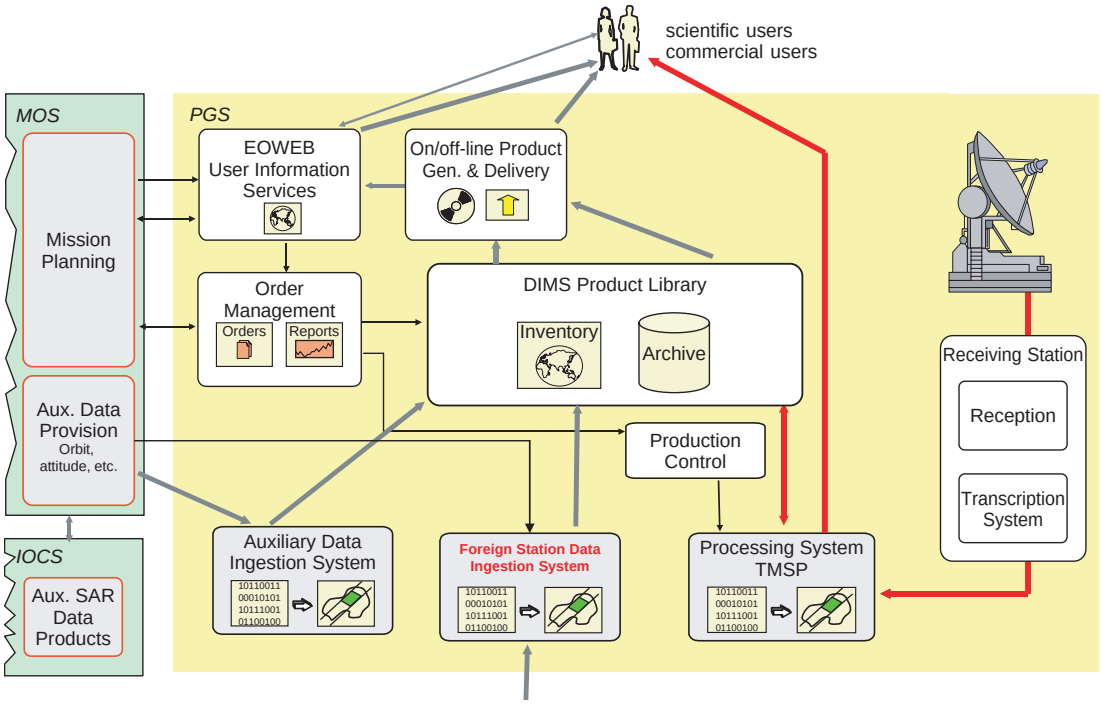


Figure 7.1.8: *Structure of a multimission ground segment (Source: DLR).*

In the past 15 to 20 years, various institutions have established and operated payload ground segments which originally followed a "single mission concept" (SPOT, ENVISAT, IKONOS). The increasing number of different types of missions and the urgent necessity to be "interoperable" made explicit the need to reduce the resources required for customized approaches as well as to set up and head a Europe-wide decentralized ground segment and encouraged the development of so-called **multimission ground segments**. Based on developments in recent years, DLR established such a ground segment for national missions such as Terra-SAR-X and EnMAP, which at the same time functions as a key element for ESA missions (Figure 7.1.8). The payload ground station aspects of data reception, processing, archiving, access and distribution are described in somewhat more detail below.

Data Reception with Station Networks

Modern Earth observation sensors generate data rates and volumes which exceed by far the bandwidth available for transmission (X-band, 8.025–8.4 GHz). The sensor data is therefore **temporarily stored, compressed** and relayed to Earth at **reduced data rates**. The data volume, the limited contact time available and the desired low latency period between data reception and data relay to the ground stations together pose a serious bottleneck for the effective use of Earth observation data. The establishment of **ground station networks** was therefore an early priority for Earth observation missions. Today a distinction can be made between two fundamental ground station profiles. The real-time-oriented ground stations, also called **direct access stations**, of the type designed for the regional operations centers, called ROCs, of the IKONOS satellites have the task of assuring immediate data access in the regions where they are located and making possible optimized data processing and applications. By contrast, **polar data-dump stations**, because of their favorable location at both poles and accordingly frequently overpasses, have the task of assuring the transfer of globally received and

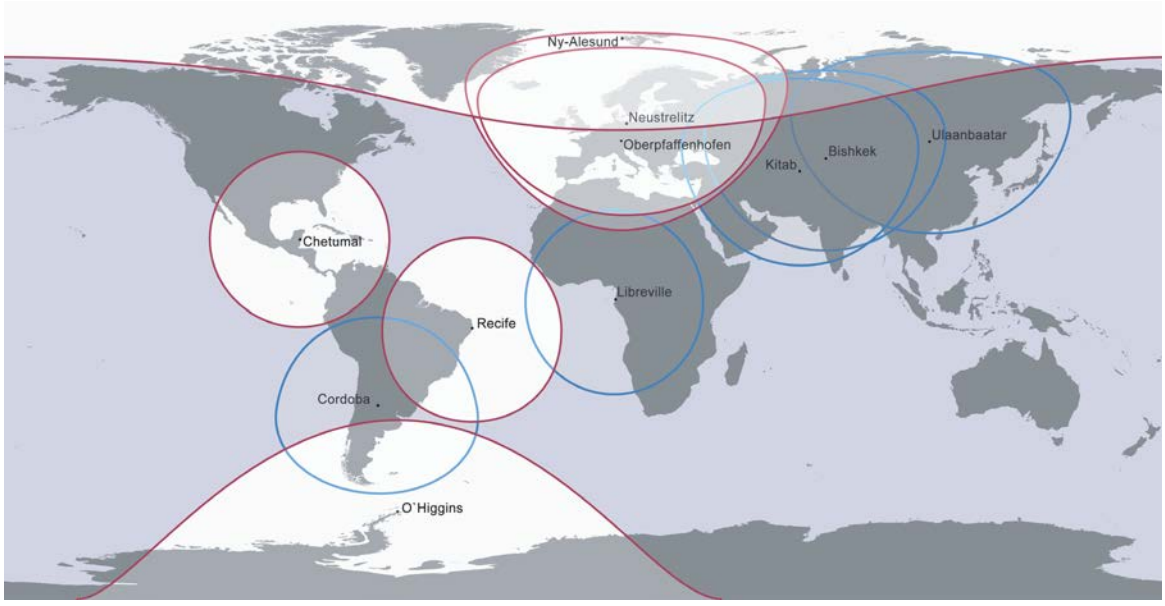


Figure 7.1.9: *DFD's station network (Source: DLR).*

stored data to central data centers using terrestrial or satellite-supported communication channels. The coordination of internationally distributed station networks is therefore one of the challenges in carrying out Earth observation missions. DLR currently operates a station network to support national and European missions (Figure 7.1.9), suitably adapted to specific mission requirements with the help of additional partner stations.

An innovative concept already tested in a few cases is the future relay of data from Earth observation satellites to a **geostationary communications satellite** via a rapid microwave link (ENVISAT) or an optical communication link (planned for the TanDEM-X mission). Contact times between the Earth observation instrument and the data center can thus be significantly increased, which makes not only large data volumes available with shorter latency periods, but also secures the transmission path from interference.

Data Processing, Archiving and User Access

The bridging function between data reception and user requirements for data and information products is comprehensively filled by the data management function of the payload ground segment. It incorporates the following services:

- **Processing** to value-added levels.
- **Archiving and cataloging** the received raw data and derived information products.
- **Order management** and distribution of the requested Earth observation products to users.
- **Online access via the Internet** to existing data sets as well as to request the recording of new data sets.
- **Monitoring and control** processes, which ordinarily operate automatically and reliably.

Data Libraries

The heart of a payload ground segment is its **archiving and cataloging system**. It stores all received raw data and derived value-added products for long-term preservation, manages this data and provides easy access to catalog information and data sets. The catalog information, which describes each individual product, is managed in databases to enable efficient searches and rapid access to millions of separate information items. The image information, by contrast, is stored



Figure 7.1.10: *DFD's robot-driven data archive (Source: DLR).*

in automated archives consisting of hard-disk systems for near-real-time access and in robot-controlled background magnetic tape archives. This staged procedure assures that there are no delays in near-real-time access when the magnetic tapes are accessed (Figure 7.1.10).

Processing

Processing systems are used within the payload ground segment to convert raw data to **value-added information products**. Depending on the instrument (SAR, multispectral or hyperspectral sensor) and the desired processing level, specific processing algorithms are used on the assigned computer platforms. These so-called (sensor) processors are integrated in the information technology environment of the multi-mission ground segment via modules which assure the uniform and harmonized management of the processors. Different categories of processing sequences regulate the data flow in the ground segment. Thus in the **ingestion** processing sequence the raw data from in-house or external ground stations is recorded and associated metadata extracted as required by the archive, for example by determining image corner coordinates, extent of cloud cover, or by deriving typical processing parameters such as the pulse repetition frequency (PRF) of the SAR sensor. Independently, **systematic data-driven processing sequences** for individual data sets can be automatically converted without external request into end products by means of predefined parameters, or archived data sets covering several years can be reprocessed on the basis of new algorithms, or refreshed at defined intervals by being copied onto new media. These systematic data flows can be contrasted with **user-driven processing sequences**. Here, individual requests for data reception are initiated and customized information products generated. Figure 7.1.11 is an orthogonal depiction of the horizontal data-driven and the vertical user-driven data flows.

User Information Services

Another important function of the ground segment is providing **online access services** for users. It permits access to the product and data catalog and to supporting functions (guides), usually via a Web portal. Searches, coupled with visualization elements, provide easy access to browse images, to detailed parameters or permit the transfer of entire scenes. At the same time the order management system initiates the order and delivery process and confirms ordering rights and user profiles.

Product Distribution

The distribution of data and information products to a wide range of users (from individual scientists to organizations which operate their own service and distribution facilities) reflects the variety of requirements which must be met by the data distribution system. Depending on the data volume and specified delivery time, data sets can be copied onto various types of media (CD, DVD, magnetic tapes or disks) for delivery. **Complex distribution services** allow the supply of individual products, product bundles from various missions, or even mass supply of data sets covering several years of monitoring. The distribution of data sets in **near real time** via terrestrial or satellite-supported networks is gaining in importance, especially for civil protection purposes and in catastrophic situations. The establishment of near-real-time delivery mechanisms as well as the generation of user-specific "postprocessing" products, such as time series, mosaics and geoinformation in an OGC-compatible format, are the goals driving current developments in data distribution services.

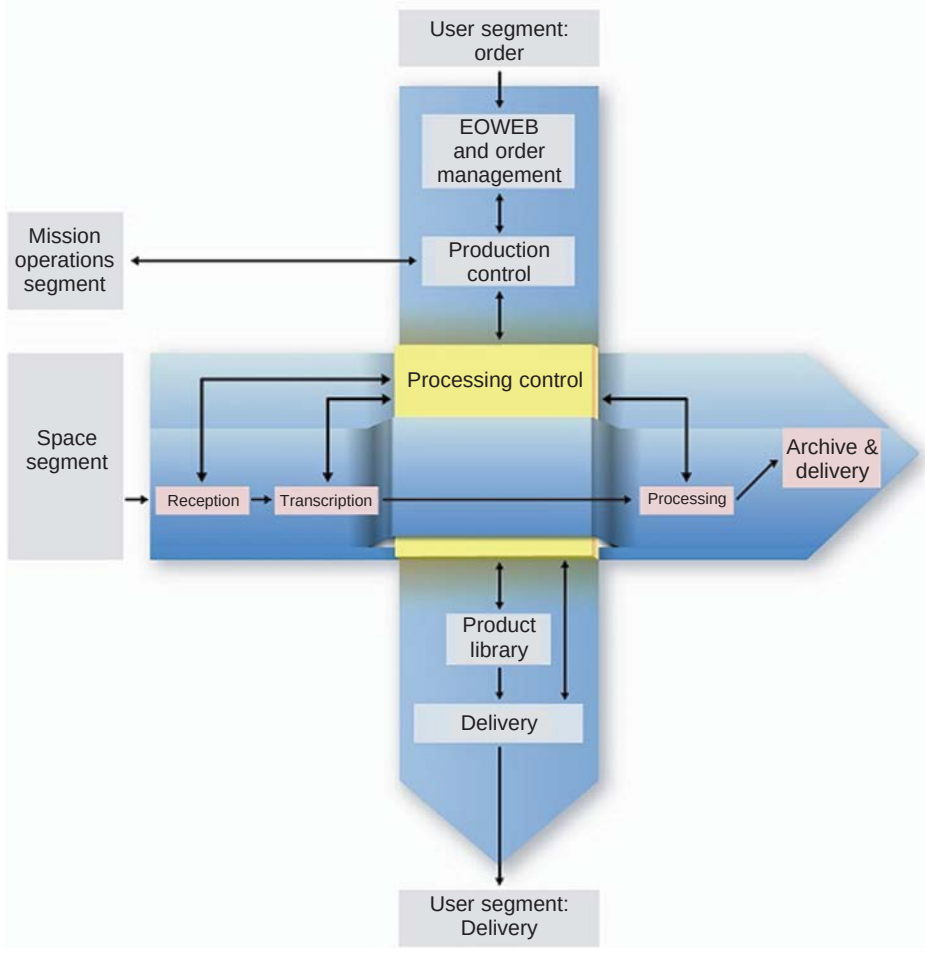


Figure 7.1.11: *Data (horizontal) and request (vertical) workflows.*

7

7.1.3 Utilization Programs and Important Earth Observation Missions

Together with the Galileo satellite navigation system, the European GMES program (**Global Monitoring of Environment and Security program**) contributes to European independence and systems leadership in future space technologies. These new Earth observation technologies will bring benefits especially for the European population in the form of new applications.

Starting in 2011, ESA's fleet of GMES satellites (SENTINELS) will continue the successful series of ESA and EUMETSAT satellites with an emphasis on environmental, weather and climate monitoring. These European systems will be supplemented by national satellites which focus on mapping and security tasks. Examples are the German satellites TerraSAR-X, RapidEye, TanDEM-X and EnMAP (see also Table 7.1.2).

Data from these satellites will provide inputs for **geoinformation services** which help us to make better use of our living space, mitigate catastrophes, enable timely recognition of environmental damage and protect us from security risks.

Digital maps, permanent monitoring of ground reference measurement equipment and aerial images

Table 7.1.2: *Survey of current satellite missions.*

1	2 Satellite or mission/sensor	3 Agency or company/ country	4 Resolution min/max in m	5 Swath min/max in km	6 # of bands/ SAR band	7 2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Remarks
	Landsat class, high-resolution/multispectral															
P	Landsat -5	NASA/US	30/120	185	7											
P	Landsat -7	USGS/US	15/30	185	8											
	Landsat Continuity	NASA/USGS	tbd	tbd												
P	Terra/ASTER	NASA,JAXA/US,JAP	15/90	60	14											Stereo
C	SPOT 1,2, 4	SPOT,CNES/F	10/20	60	5											3 satellites
C	SPOT 5	SPOT, CNES/F	2.5/5/10	60/120	5											
C	AstroTerra	SPOT, CNES/F	2/8	60	Pan/4											
	SEOSAT	CDTI, INTA/ES	2,5/10	30/60	Pan/4											
	Sentinel-2	ESA/EU	10/30 tbd	200/400	8 + (tbd)							1	2			2 satellites
	DMC	SSTL./UK&others	32	640	3											3 satellites
	DMC BilSat-1	SSTL/Turkey	12/16	25/55	3											
	DMC+4 CHINA	SSTL/China	4/32	24/600	3											
C	DMC	Deimos, LATUV/ES	22	600	3											
	SAC-C/MMRS	CONAE/Argentina	175	360	5											
C	ResourceSat-1,2	ISRO, ANTRIX/IND	5.8/23	23.5/140	3/4		2									2 satellites
	CBERS-1,2	CAST/CHIN, INPE/BRA	20/80	120	8											2 satellites
	CBERS-3	CAST/CHIN, INPE/BRA	5/10/80	60/120	4/8											
P	ALOS/PRISM	JAXA/JAP	2.5/10	35/70	5											Stereo
C	RapidEye	RapidEye/D	6.5	78	5											5 satellites
	Lower resolution multi/hyperspectral															
P	Terra,Aqua/Modis	NASA/US	250/1000	2330	36											
C	Envisat/MERIS	ESA/EU	300/1200	575/1150	15											
	EO-1	NASA/US	30	15	233											
C	ResourceSat-1,2	ANTRIX, ISRO/IND	70	720	3		2									2 satellites
	EnMAP	DLR	30	30	>200											
	Very High Resolution Optical															
C	IKONOS-2	Space Imaging./US	0.8/3.2	11	1/4											
C	EROS-A1	ImageSat/ISRAEL	1.8	13.5	pan											
C	EROS-B1-3	ImageSat/ISRAEL	0.7/2.8	10.4	1/3	1	2	3								3 satellites
C	Quickbird-2	DigitalGlobe/US	0.6/2.5	16.5	1/4											
D	WorldView-1	DigitalGlobe/US	0.5	16	pan											NextViewI
D	WorldView-2	DigitalGlobe/US	0.5/1.8	16.5	1/8											
C	OrbView-3	Orbimage/US	1/4	8	1/4											
C	GeoEye-1	Orbimage/US	0.4/1.6	15.2	1/4											NextViewII
C	FormoSat-2	NSPO/Taiwan	2/8	24	1/4											Stereo
	ALSAT-2	ASAL, CNTS/Algeria	2,5/10	17,5	1/4			?	?							2 satellites
	THEOS	GISTDA/Thailand	2/15	22/90	1/4											Stereo
C	CartoSat-1	ISRO, ANTRIX/IND	2.5	30/60	pan											Stereo
C	CartoSat-2	ISRO, ANTRIX/IND	1	12	pan											
	KOMPSAT-2/3/6	KARI/Korea	1/4 (2,8)	15	1/4		1			2					3	3 satellites
	Resurs-DK1	Spaceproject, FSA/RUS	1/ 2.5	28.3	1/3											
	TOPSAT	BNSC/MoD	2,5/5	25	1/3											
D	Pleiades-1/-2	CNES/F, SISA/F	0.7/2.8	20	1/4					1	2					2 satellites
	Synthetic Aperture Radar (SAR)															
P	Radarsat-1	CSA,RSI/CAN	8/100	45/500	C											
C	ERS-2	ESA/EU	30	100	C											
C	ENVISAT/ASAR	ESA/EU	30/150	100/400	C q											
P	ALOS/PalSAR	NASDA/JAP	7/100	70/350	L Q											
P	TerraSAR-X	DLR & InfoTerra/D	1/16	10/100	X Q					X						&TanDEM-X
	Sentinel-1	ESA/EU	5/80	240/400	C Q							1	2			2 satellites
D	COSMOSkyMed	ASI/I	1/100	10/200	X q			1	2	3	4		X			&SABRINA
C	Radarsat-2	RSI/CAN	3/100	10/500	C Q											
	SAOCOM 1A/B	CONAE/Argentina	10/100	30/320	L q					1	2					2 satellites
	RISAT	ISRO/IND	3/50	10/240	C											
D	KOMPSAT-5	KARI/Korea	1/tbc	tbc	X (tbc)											
D	TecSAR	Israel	1/tbc		X											

Abbreviations:

Column	Explanation
1	C = commercial mission or commercialized outside country of origin; P = public/private partnership between space agency and commercial partner; D = dual use: military and commercial; no sign indicates limited regional or science availability
6	Capital letters C, L, X denote radar frequency bands; Q = fully polarized; q = partially polarized; pan = optical panchromatic only
7	Numbers denote launch of satellite in a series; mission end dates are as planned or estimated; "--" denotes that satellite might operate longer
8	Numbers of satellites in the mission; special capabilities are abbreviated

contribute to a better understanding of the data coming from Earth observation satellites. Taken together they provide information necessary for decision makers in the realms of politics, society and economics.

The GMES pilot services are to supply reliable information relating to the land surface, the oceans, natural risks, security and the atmosphere starting in 2008. Initially, the establishment of these services will be jointly financed by the European Commission and ESA.

In the long term, however, GMES will be carried by the needs of the public, government authorities and industry for information about our environment and the preservation of our security.

Bibliography

[7.1.1] Webb, A. Corona: America's First Spy Satellites. Top-secret reconnaissance mission shrouded in secrecy for 19 years under guise of medical research. *Public Aff. Space Missile Times*, **14** (33), 2000.

[7.1.2] Dech, S. Looking down our Earth. In Feuerbacher, B., Stoewer, H., *Utilization of Space Today and Tomorrow*. Heidelberg: Springer Verlag, 2005.

[7.1.3] Tetzlaff, A. Coal Fire Quantification Using ASTER, ETM and BIRD Satellite Instrument Data. Dissertation, LMU München, Fakultät für Geowissenschaften, 2004.

[7.1.4] Schlüter, N., Heygster, G. Remote Sensing of Antarctic Clouds with Infrared and Passive Microwave Sensors. *Meteorol. Z.*, **11** (1), 21–36, 2002.

[7.1.5] Czekala, H., Simmer, C. On Precipitation Induced Polarization of Microwave Radiation Measured from Space. *Meteorol. Z.*, **11** (1), 49–60, 2002.

[7.1.6] Hong, G., Heygster, G., Mikoa, J. *et al.* Sensitivity of Microwave Brightness Temperatures to Hydrometeors in Tropical Deep Convective Cloud System at 89–190 GHz. *Radio Sci.*, **40**, RS4003, 2005.

[7.1.7] TRMM Propical Rainfall Mission, Instrument Description. www.nasa.gov, 2004.

[7.1.8] Ingmann, P., Lajas, D. *ADM-Aeolus ESA's Doppler Wind Lidar Mission*. Noordwijk: ESA, 2007.

[7.1.9] Ingmann, P., Straume, A.G., Lajas, D. *et al.* and Members of the MAG ESA-ESTEC, EOP-SMA/TEC-EEP. ADM-Aeolus: ESA's Doppler Wind Lidar Mission. Third Symposium on LIDAR Atmospheric Applications, 2007.

[7.1.10] Brieß, K. *et al.* Technology Demonstration by the BIRD Mission. www.dlr.de, 2001.

[7.1.11] Feuerbacher, B., Stoewer, H. (eds.) *Utilisation of Space*. Heidelberg: Springer Verlag, 2006.

[7.1.12] Schreier, G., Dech, S. High Resolution Earth Observation Satellites and Services in the Next Decade – A European Perspective. *Acta Astronaut.*, **57** (2–8), S.520–S.533, 2005.

[7.1.13] Forcada, E., Kohlhammer, G., Casgrain, C. *et al.* European Earth Observation Ground Segment Coordination, International Astronautical Congress, Valencia, October 2006.

[7.1.14] Kramer, H. *Observation of the Earth and its Environment*, Fourth Edition. Heidelberg: Springer Verlag, 2002.

[7.1.15] CEOS Earth Observation Handbook. www.eohandbook.com, 2007.

7.2 Communications

Hans Dodel

In a world in which terrestrial communications infrastructures were still in their initial stage even in Europe, the use of space for communications in the first 30 years (1957 to 1987) was quite beneficial; it had considerable practical and economic advantages for the users, and it then became the most lucrative commercial space application.

Ten years after the advent of fixed satellite communications and the first relaying of television news, satellite mobile telephony began, initially in the maritime sector, then in the aeronautical realm and on land, whereby land applications, with their considerably higher number of land vehicles compared to ships and aircraft, rapidly outdid the other two in numbers of subscribers.

With time, terrestrial systems matured, including the transoceanic cables. At the same time people exhibited a growing desire for a large number of television programs, so that satellites rapidly replaced the existing terrestrial transmitters and cable facilities and their quite limited numbers of program channels.

In addition to communications and television broadcasting, global satellite position determination was an application directly derived from the first artificial satellite, Sputnik 1, and led to the global satellite positioning systems of Zykade (Soviet Union/Russia) and Transit (USA) with Doppler measurements. Since

then, one-way ranging technology has been employed (see Section 7.3).

7.2.1 The Beginning – Sputnik

Until 1963 satellite technology was exclusively a secretive military matter, in both the Soviet Union and the USA. Then President John F. Kennedy ordered in the Communications Act of 1963 that it be employed for the benefit and welfare of humankind (an idea anchored in the US Constitution, which nobly declares an inalienable right to the "pursuit of happiness").

In contrast to the usual political rhetoric, where things may be promised but left unrealized, President Kennedy ensured that his Act was implemented. He founded COMSAT and commanded that an international operating agency be established and charged with offering satellite communications worldwide, also to those nations not yet able to pay for it. There was a clear aim to found a company which would arrange for the establishment of this international operator.

7.2.1.1 The Entry – COMSAT

Founded in 1963, the commercial company Communications Satellite Corporation, Washington, DC, received initial financing of $300 million from the US government. With this, it was able, within 10 years, to establish the independent International Telecommunications Satellite Consortium (INTELSAT); it continues to offer communication services at reasonable cost worldwide. It is also a company listed on the stock exchange, and it is the world's largest satellite operator in terms of number of satellites in orbit (53 operational spacecraft in 2008).

7.2.1.2 Mobile Communications – MARISAT

In 1973, COMSAT took the second step and established maritime satellite communications. The first MARISAT satellite was launched 30 months after the program started, and in 1976 the worldwide MARISAT satellite system for maritime communications was put into operation. For its international operations, COMSAT then established a second international consortium, the International Maritime Satellite Organization (INMARSAT, London). In 1983, INMARSAT took over the satellites and their operation from COMSAT, a task it continues to carry out independently and successfully today. It has since included aeronautical and especially terrestrial mobile communications.

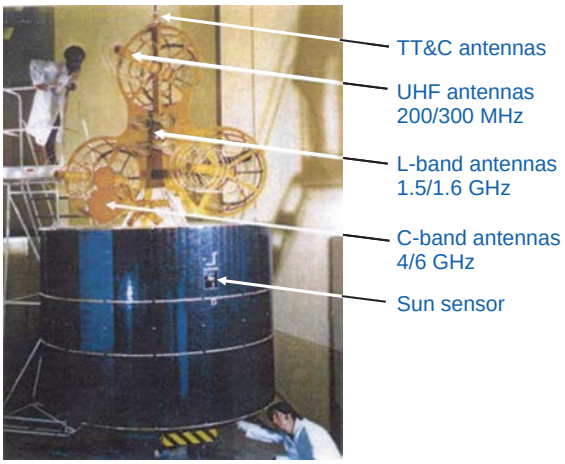


Figure 7.2.1: *The MARISAT satellite with its UHF and L-band antennas on a spin-stabilized cylindrical body (the axis of rotation is parallel to the Earth's axis) and despun Earth-oriented antenna (Source: COMSAT).*

Figure 7.2.1 shows the MARISAT satellite, a spin-stabilized satellite (the axis of rotation is parallel to the Earth's axis) with a despun Earth-oriented antenna.

7.2.1.3 The National MOLNIYA Program

During this time the Soviet Union achieved optimal coverage of its vast territory extending over 11 time zones with its nongeostationary MOLNIYA satellites. These satellites also served "brother" nations such as Cuba, North Vietnam, the GDR, etc. The disadvantage of these satellites, namely that numerous ground antennas must track the satellites 24 hours a day, was not trivial, since, besides those run by the state, there were few private ground stations. This satellite system was, despite the founding of the INTERSPUTNIK organization in Moscow and its mandate to offer satellite transmission free of charge to communist "brother" nations, scarcely of any use for the civilian population, quite contrary to the humanitarian call of John F. Kennedy.

In contrast to INTELSAT, INTERSPUTNIK did not own any satellites but was allocated communication capacity according to need by the Soviet government – some 300 telephone circuits worldwide in 1990 (compared to 300 000 telephone circuits and over 100 television channels for INTELSAT).

Later, the GORIZONT fleet of geostationary satellites was built up with greater capacity (in the C-band). This introduced a transition from global systems to regional and national satellite telecommunications networks.

7.2.1.4 National Satellite Networks

One of the first national civil networks, besides the military networks of the USA and the Soviet Union, was established in 1976 in Indonesia (a country of 14 000 islands), and almost simultaneously in Canada with the ANIK system (extending over 5000 km of the sparsely settled northern part of the American continent).

At that time, ARABSAT started broadcasts for the Arab language area, offering also telephony and television services.

Then, gradually, national systems were set up, first in the USA. The RCA SATCOM had the most modern satellites of the time, also serving Alaska. Other regional systems followed for particular zones of interest. The EUTELSAT (European Telecommunications Satellite Organization) was created to supply coverage in Europe.

It should be emphasized that at a time in which national terrestrial telephone systems were still analog and thus of limited transmission capacity (with data rates of 64 kbit/s at most) these satellite system offered megabits per second services between, say, Scandinavia and the Iberian Peninsula.

The failure of the German television satellite TV-Sat provided the Société Européenne de Satellits (SES, Luxembourg) with an opportunity to offer pan-European television coverage with the ASTRA system, a service which EUTELSAT joined later. Today it is difficult to imagine life in Europe without satellite television transmission.

Next to television, Europe-wide satellite radio broadcast is growing, starting with the WORLDSPACE system. Radio, in contrast to television, is of course much more determined by regional cultural interests: one is more likely to view a Hollywood film than, say, listen to Portuguese songs in Scandinavia, or vice versa. The essential attribute of satellites, to be able to serve large areas, is not demanded here; the radio stations with the most local appeal for listeners appear to be the most successful.

7.2.2 Satellite Communication Services

Worldwide, there are over 100 satellite operators in the communications business offering telephone services and data transmission, mobile telephony, television and radio [7.2.1], specifically in the fields of:

- Fixed satellite services (FSS)
- Mobile satellite services (MSS)
- Broadcast satellite services (BSS).

These three categories, out of a total of 20 satellite transmission services (see Table 6.3.2), are the most profitable.

Since the beginning of commercial space flight in 1965 with Intelsat 1, the vast majority of satellites fly in **geostationary Earth orbit** (GEO) (see Section 2.2). The official designation for this orbit is the geostationary orbit (GSO), and for all other orbits the nongeostationary orbit (NGSO).

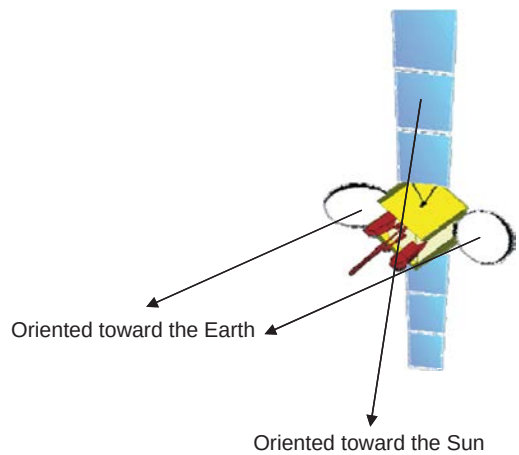


Figure 7.2.2: *A communications satellite containing three flywheels to stabilize its position in the three coordinate axes [7.2.3].*

These satellites have their antennas oriented toward the Earth while they are orbiting, but their **solar panels** always need to be oriented perpendicular to the Sun in order to receive the maximal amount of solar energy (see Figure 7.2.2). For this purpose they are mounted on the *z*-axis so that they can rotate around their axis once a day relative to the body of the satellite (the satellite's *z*-axis is parallel to the Earth's axis). The electrical power generated is transferred to the satellite body via slip-ring bearing and the power transfer assembly (BAPTA).

The antennas have to be aligned with the target area with a precision of 0.1° (and in the future the precision specification will be even more stringent as antenna sizes and particularly frequencies increase).

7.2.2.1 MOLNIYA Satellites and Orbits

The exception to the rule that commercially successful satellite systems fly in GEO are the Russian MOLNIYA satellites. In order to serve this massive country, at least three GEO satellites are required (with only two, the north–central region would lack coverage). The highly elliptically orbiting MOLNIYA satellites (they bear the same name as their orbits and the rockets used to launch them) fly in a relatively stable orbit inclined 63.43° relative to the equatorial plane. The MOLNIYA orbits are so designed that the satellites stay over Russia for a long time (apogee), with only a short flight over the South Pole (perigee). Over northern Siberia, the MOLNIYA satellites offer an almost vertical view angle (instead of a grazing angle in the case of GEO satellites, if they can be seen at all).

A MOLNIYA orbit is used for satellite-supported pay radio in the USA, the SIRIUS system. In a 24-hour orbit with only two satellites, one is always over the USA during its long time close to apogee.

7.2.2.2 Local Services and the Last Mile

Other national satellites are used occasionally in the USA for very local applications, such as for vacationers in isolated national parks so they do not have to do without the Internet there.

The most exotic topological, but nevertheless very up-to-date application for geostationary satellites (whose essential attribute is coverage of almost half of the Earth) is the link for the last mile from the street cabinet in the neighborhood to the home. Particularly in North America, where this last mile can be much longer and with users today expecting gigabit per second services, a satellite connection is faster to install and less expensive than a cable connection for this bit rate [7.2.4].

7.2.3 Low-Orbit Spacecraft

Satellites flying in low Earth orbit (LEO) were never used commercially, even in the early days of space technology. The reasons are that a large number of satellites are needed to achieve geographical coverage, and at each ground station two to three antennas are necessary to provide uninterrupted service (one antenna tracks the descending satellite, e.g., in the west; the second looks toward the ascending satellite in the east, while a third is "scanning" the sky for a more suitable satellite, possibly at a higher altitude, less blocked by buildings or with more free capacity).

The supposition that launch into low Earth orbit is less expensive is a fallacy. Placing satellites into orbit at the geostationary altitude of 36 000 km above the equator has not been an obstacle ever since Harold Rosen of the Hughes Aircraft Company, USA, developed the apogee kick motor (AKM) and the associated launch procedure in 1965 and successfully demonstrated them with Intelsat 1. With the AKM it is possible to put satellites into GEO with rockets limited to LEO. The satellite's internal propulsion system, even today, remains the most efficient method of taking a satellite to GEO altitude. It competes with transport via GEO-capable rockets, which have the substantial disadvantage that they themselves remain in the geostationary orbit where they could collide with satellites.

Whereas communications satellites fly as high as possible, preferably in geostationary orbits, so that they can serve the largest possible region (42.6% of the Earth), Earth observation and reconnaissance satellites fly in low orbits to maximize the resolution of their imagery. An exception to this rule was proposed during the Cold War as an answer to the question of how to make a (military) satellite invulnerable. One single satellite cannot be made invulnerable, but many satellites can be launched and operated so that it is

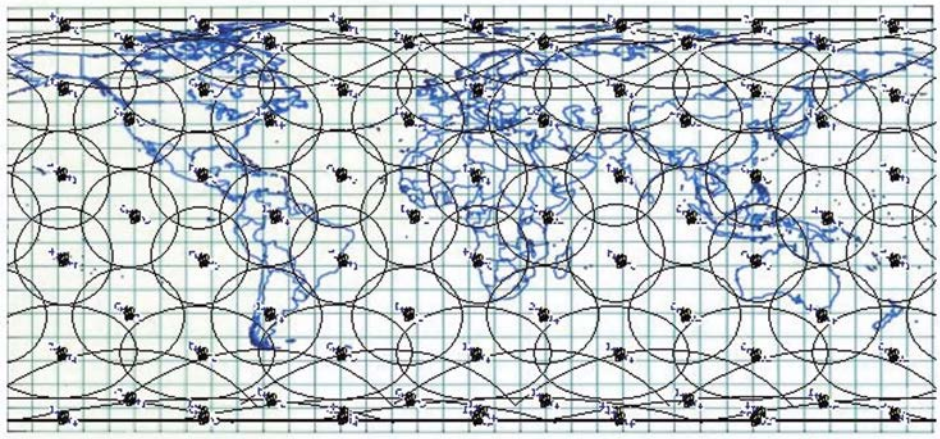


Figure 7.2.3: *The footprints of the Iridium satellite constellation.*

technically and especially economically too costly for an opponent to disable all of them.

This idea led to the concept of hundreds of low-flying satellites (the Star Wars/Brilliant Pebbles program). After 1990, the number of satellites was reduced to only 77, later to 66, still a large number, particularly since the requirement of invincibility of the fleet was no longer a systems requirement.

Nevertheless, the Iridium, Globalstar and OrbComm communications systems were realized. They joined the intermediate altitude circular orbit (ICO) defined by INMARSAT. All four companies predictably went bankrupt in 2000, from which Iridium, Globalstar and OrbComm recovered with the help of US government contracts. The commercial success of low-flying communications satellites has not been proven to date.

Figure 7.2.3 shows the footprints of the Iridium satellite constellation of low-flying satellites. Iridium flies in a polar orbit at an altitude of 790 km and has its highest satellite density at the Earth's poles. Over the world's most heavily populated areas at about 45°N latitude, only a single satellite is in view (whose failure leads to service outage).

Figure 7.2.4 shows the footprints of the Globalstar constellation of orbiting satellites. In contrast to Iridium, the Globalstar satellites fly at 1414 km altitude in inclined orbits and have their highest concentration over the latitudes of the Earth's population. Here up to four satellites are in view, so that the failure of any one of them will not interrupt service.

The advantages and disadvantages of satellites in GEO and LEO are summarized below based on various considerations.

Launch Cost: About 50% of the cost of launch into GEO arises during the first 10 km. The higher unit cost for a GEO is outweighed by the large number of satellites necessary in LEO. Low orbits have no advantage over geostationary orbits as to their energy balance.

Transmit Power: Recalling the law of the conservation of energy, the power received on the Earth's surface is independent of the flight altitude from which it is transmitted. So LEO offers no signal advantage compared to GEO.

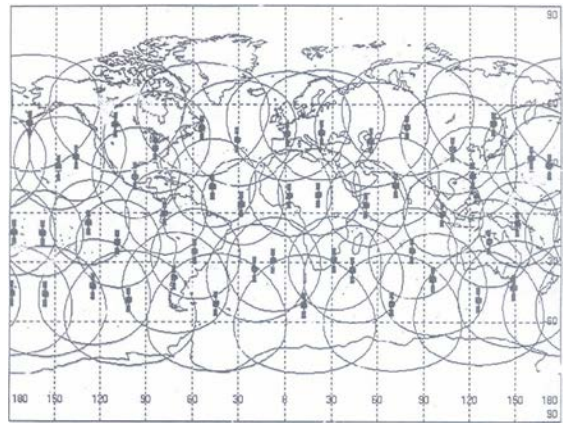


Figure 7.2.4: *The footprints of the Globalstar satellites.*

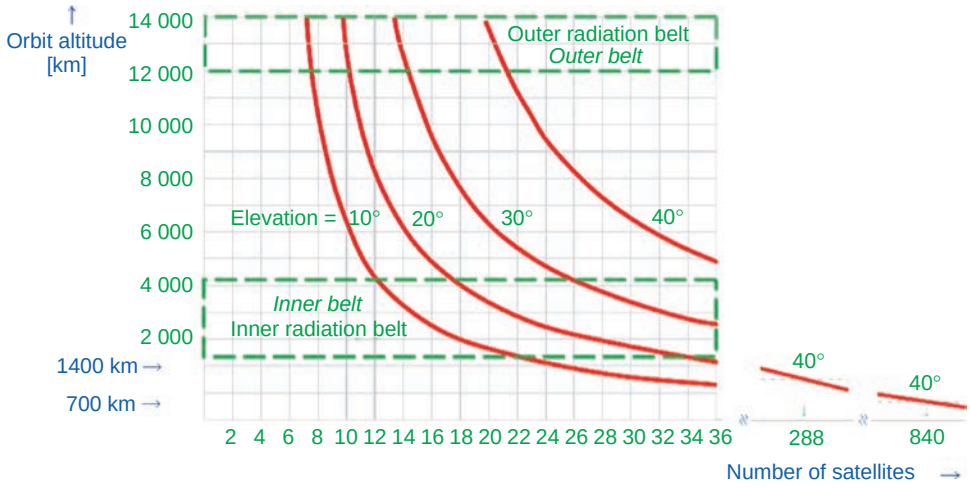


Figure 7.2.5: *The number of satellites required for coverage as a function of elevation angle and flight altitude.*

Frequency Utilization: The efficient use of the limited frequency spectrum leads to the use of individual cells on the ground, where a frequency band can be reused in every other cell; so the smaller the cells, the higher the exploitation of the spectrum. The cells projected on the Earth's surface from the geostationary Inmarsat 4 satellite are as small as those of low-flying Globalstar, so the low orbiters have no systems advantage here, either.

Signal Latency: Defined as the time needed for a communication signal to travel to the satellite and back, system latency has often been listed as an advantage of the low orbiters. This argument was not always correct either. For satellites having on-board signal processing, there can be latencies as high as the time required for the signal to reach GEO because of the processing time.

For broadcasting services (television) low orbiting satellites are totally inappropriate because the users' antennas would have to track those satellites 24 hours a day.

In the case of other services, particularly Earth observation and reconnaissance, only a small number of ground stations are operated and it is desirable to fly as low as possible. But such services do not require global 24-hour coverage; these satellites fly frequently but not continuously over a particular target area.

Today, communications services have to be globally and continuously available to be successful in the market. If low orbits are selected, then a large number of satellites have to be launched and operated. Figure 7.2.5 shows the number of satellites required for a given minimum elevation between the user and the satellite as a function of the altitude of the satellite (see also Section 2.2).

The Figure illustrates, for example, the case when a satellite flies at an altitude of only 700 km with a user requiring a minimum elevation angle of at least 40° so that there is a good chance of seeing the satellite also in urban environments. This leads to a requirement for 840 satellites, the original TeleDesic system, which could not be financed.

The horizontal bars in Figure 7.2.5 represent the inner and outer Van Allen radiation belts, which should be avoided for technical reasons related to on-board electronic equipment. Reasonable flight altitudes are under 2000 km (the LEOs), between 6000 and 10 000 km (the ICOs) and above the outer Van Allen belt (such as at 20 000 km for navigation satellites).

7.2.4 Satellites in Medium-Altitude Orbits

The need to reduce the large number of satellites required for the LEO systems led to the proposal of systems with constellations in ICOs proposed by ICO Inc. and TRW Corp. (the Odyssey system). The ICOs are about 10 000 km high (instead of 700 to 1400 km

for LEO constellations). With two orbital planes of five satellites each, or three planes with three satellites each, global coverage is achieved, including the Earth's poles. The time for the signal to travel forth and back is almost negligible at 70 ms (assuming that no on-board signal processing takes place in the satellites).

But as in the case of the LEO, the user antennas, instead of remaining stationary, have to constantly track the orbiting satellites, with the additional expense of a second antenna per terminal pointed toward the ascending satellite while the first one operates with the descending satellite, for (almost) uninterrupted service.

An additional requirement is signal handover from the descending to the ascending satellite, several times a day, with the smallest possible loss.

For general communication services and particularly for the broadcast of television programs, ICO orbits are not suitable.

7.2.5 Satellites in High Orbits

In higher orbits, twice the height of ICOs, satellites are also moving in circular medium-altitude Earth orbits (MEOs). The GPS/NAVSTAR system of the US armed forces and the GLONASS system of the Russian armed forces both fly at 20 000 km altitude, and Europe's Galileo at about 24 000 km [7.2.5].

At this flight altitude the diameter of the ground cell subtended by the satellite is about 20 000 km, just less than 40% of the Earth's surface. A minimum total of six satellites would thus suffice to have at least one of them always available for every point on Earth. With 18 satellites at least three would always be in sight, so that triangulation could be carried out to determine position.

There are no known projects for communication systems in MEO. Such systems would have the same difficulties relating to the need for tracking with at least two antennas per ground station, including the handover of the signal, as described for the LEO and ICO orbits [7.2.6].

7.2.6 Satellites in Highly Inclined Orbits

The highly inclined elliptical orbits (HEOs) (often incorrectly interpreted as highly elliptical orbits) are in principle suitable for communications, the most well-known examples being the MOLNIYAs mentioned above for telephony and data (Russia, since 1965) and the SIRIUS system for radio/music (USA, since 2000).

The high inclination permits coverage of regions in the high north (see Figure 7.2.6); in culmination, MOLNIYA sees not only the North Pole, but also the other side of the Earth down to New York.

The 63.43° inclination guarantees a stable orbit with minimal requirements for operational adjustment. With a 24-hour orbit, two satellites equally distributed in orbit are sufficient for one of them to always be over the target area.

HEOs can be operated for any orbital period. With a 12-hour orbit, two northern culminations would be achieved. Global HEO constellations with, for example, three culmination points over Europe, the USA and Japan/China (see Figure 7.2.7) could presumably serve 90% of the Earth's population. The operator of communication services based on

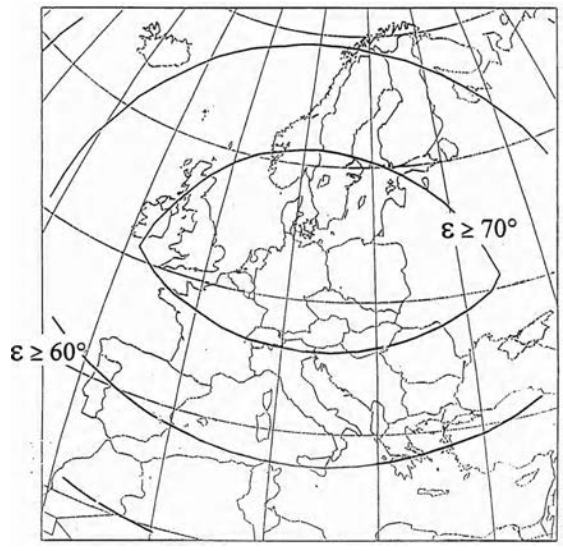


Figure 7.2.6: *The elevation angle ε to a satellite in HEO including northern latitudes.*

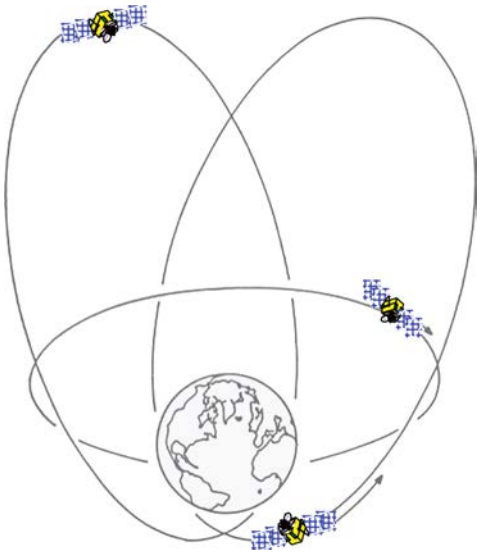


Figure 7.2.7: *Three HEOs in three orbit planes offset by 120°.*

this global system would, however, have to obtain operating licenses in North America, Europe, China and Japan (which is not a trivial undertaking). The operator would also have to set up and operate service organizations in these regions, something that could not easily be done against the will of the established telecommunications organizations there, some of which are still state owned.

7.2.7 Satellites in Inclined, Geostationary Orbits

An orbit only conditionally suitable but nevertheless very promising for communications is the inclined geosynchronous orbit (IGSO).

Geosynchronicity means that in IGSO the duration of one orbit is 23 hours, 54 minutes and 6 seconds (the sidereal day) so that the satellite always remains at the same longitude.

Inclination refers to the fact that the IGSO orbital plane is inclined with respect to the equatorial plane; over the course of a day a satellite in this rotating inclined orbital plane travels north–south relative to the Earth (just as someone on a rotating tipped ride at an amusement park goes up and down relative to the solid ground).

The length of the north–south movement amounts to twice the value of the inclination. IGSOs with a slight inclination are a common occurrence in geostationary Earth orbits: the asymmetrical attraction of the Moon relative to the equatorial plane, which affects a satellite in geostationary orbit, shifts its orbit from the horizontal plane and inclines it by 0.8° per year while maintaining the orbital period.

Uncorrected, the satellite would gradually increase its inclination up to a maximum of 17.5°. At this inclination the satellite would see the Earth's poles once daily, long enough for a communication link with research stations there.

To cover northern regions at large angles of incidence, IGSOs have to achieve higher inclinations; the satellites are then not launched from equatorial bases (such as Kourou, less than 5° latitude) but rather from northern sites (such as Plesetsk, at about 63.43°N).

With three or more satellites in IGSO, one of them is always in the northern culmination. If the longitude of the IGSO is in Europe, the satellites fly over Central Europe and provide users there with an almost vertical elevation angle, in contrast to an elevation angle under 30° which Central Europeans have with a GEO. This is an important precondition for communications in general, especially in urban environments.

These advantageous elevation angles and the regional attribute of IGSO led to the selection of IGSO in the original Galileo system planning, with the goal of beginning the system implementation in Europe. A European IGSO would not have flown over the USA; there were fears that the USA would not tolerate a foreign competitive system. Also, European coverage was considered preferable to worldwide expansion, especially for mobile land applications in Europe, which account for over 90% of the utilization of a satellite positioning system.

The extension of Europe's Galileo to achieve international coverage with IGSOs over the Americas and Asia was intended, should these regions join the system and participate financially. In this way a global system could have been built up step by step, with international cooperation and participation, including especially the USA.

Simulations also showed that an IGSO constellation would be more effective than an MEO constellation; with the same number of satellites it

achieves better coverage, or it requires fewer satellites for the same coverage. Nonetheless, MEO technology was selected for Galileo for political reasons, so the system is not sovereign; it can be degraded or turned off by the USA whenever and wherever it considers this to be necessary, with the explicit permission of the EU.

The value of IGSO for communications is in the high elevation angles, in northern regions as well, which permits communication also in urban high-rise environments with no visibility to GEO satellites. Even more relevant is the characteristic of IGSO to be able to reuse the frequency spectrum employed in GEO without causing mutual interference. The frequencies used there can be reused in the northern culmination and once again in the southern IGSO culmination, a tripling of the usable spectrum; the orbital communications capacity is tripled with IGSO.

This characteristic outweighs all other points of view. The available spectrum in GEO is so extensively used today that it is almost impossible to obtain licenses for and operate any more geostationary satellites to supply the Earth's highly populated areas.

There are presently about 500 commercial communications satellites in GEO in active operation or with legal licenses to operate. In addition, there are military communications satellites and satellites providing weather services, used for civilian Earth observation or employed in military reconnaissance.

For higher data rates, Internet downloads and broadband video from unmanned flight drones, all of which are transmitted to the control centers via geostationary satellites, larger ground antennas are required.

Compared to GEO satellites, satellites in IGSO have to operate with ground antennas which are either small enough for the satellite to be in the antenna's main beam, or have larger apertures, which must be dynamically tracked. However, in the latter case a single antenna is sufficient for each ground terminal (in contrast to the three antennas required for LEOs). Since the ascending and descending satellites are at the same longitude, they meet at the point of signal transfer and a simple handover at this location is possible.

Today, modern antenna designs are available for antennas with large apertures whose beams can track satellites. Instead of parabolic dishes of sheet metal, an arrangement of patches is used, which can be controlled in amplitude and phase so that they generate a beam in the target direction, and this without employing any mechanically moving parts; these are the so-called phased array antennas.

The directivity and thus the antenna gain, which are important for the transmission, are identical for phased array antennas of the same size as parabolic dishes operating at the same frequency. Since the phased array antenna consists of semiconductor elements, it has the potential for significant cost reductions when produced in sufficient quantities.

7.2.8 Polar Orbiting Satellites

Satellites in orbits crossing over the poles are suitable for Earth observation, not because of their view of the poles, but because polar Earth orbits (PEOs) can be so designed that after a few orbits the satellite repeats its track; it flies with the same Sun angle so that new image material can be compared with previously recorded images without misleading shadows.

For communication services (stationary or mobile telecommunications and television), polar Earth orbits are unsuitable and economically undesirable because they produce the highest satellite density (and thus systems costs) at the poles where there are no customers. In contrast the satellite density is relative low, where the higher user density (population density) is located on the Earth's surface.

7.2.9 Platforms in the Stratosphere

Long before space activities began, high-flying platforms were used for communications and especially for television broadcasts (Ohio, USA, 1955). Today, there are still special applications, so a glance at these stratospheric satellites is appropriate. The radiofrequency regulations today provide frequency allocations for communication with high-altitude platforms (HAPs).

Aircraft and zeppelins are used as high-altitude platforms, less so balloons, which, unlike aircraft and zeppelins, cannot be maintained in a stationary

position against the forces of the jet streams which dominate the upper atmosphere.

The advantage of aircraft is their ability to easily maintain their position against jet streams, an essential precondition for communications services. In addition, they can be brought back down regularly and at short notice for servicing and refueling.

The advantage of zeppelins is that they can maintain altitude without fuel, but they too have to resist the jet streams. The use of fuel is almost impossible for this purpose, since it would be too costly to bring it up to the vehicle, or to bring the zeppelin back down to Earth for refueling.

Perhaps there will be, one day, affordable batteries to store enough electrical power during sunshine to maintain the zeppelin's position during nighttime.

To cover an area the size of Germany, an altitude of about 20 000 m, which is above the airspace used by aircraft, is necessary and adequate. For the coverage of larger regions a fleet of several zeppelins would have to be in operation simultaneously.

In any case, an HAP in the stratosphere would have to precisely maintain its position also during the night, to allow communication at high bit rates, and thus with large, fixed, directive, nontracking ground antennas.

If these preconditions are met, then HAPs are platforms of choice for communications, which are primarily regional (one calls more often someone in the same county than someone on the other side of the globe). Precisely this user profile can be well served with HAPs, which would free satellites from this service.

7.2.10 Telecommunication Services: Little–Big–Mega

In the context of satellite communications, the vocabulary little, big, super and mega is used; what does that mean [7.2.7]?

7.2.10.1 Little Services

Services with low and very low data rates (short message services) are also known as little services (little LEO services). These satellites typically fly in circular inclined orbits at low elevation (LEO) and store data on-board, which is demodulated so that the addresses can be read. When flying over the addressee the data is then sent down. Alternatively, with many of these little LEO services users can themselves download their data from the satellite storage system.

Even when the satellites deposit the message data immediately after identifying the address, transmission is not synchronous, so it is not suitable for voice traffic but only for data. In addition, the frequency bands used in the UHF range are so overloaded by various services that it can take up to 15 minutes before the user terminal can deposit a message; there is no real-time capability.

Besides the little LEOs for the little services, small amounts of data can also be transferred via geostationary satellite systems like EUTELSAT (EutelTracks, in the Ku-band) and via INMARSAT (Standard C and Standard D, in L-band) inexpensively, reliably and without delays.

7.2.10.2 Big Services

This type of service differs from the little services in the bit rate; big services (big LEO services) can also transmit voice and data at ISDN speed. The term is, however, restricted to low-flying satellites; the geostationary communication satellites have of course always been used to transmit voice and data.

The two most well-known big LEO systems are Globalstar and Iridium, with Iridium also operating little LEO services. The latter is a packet-switched network suitable for transmitting even the smallest data units, whereas Globalstar is circuit switched. For this, it would not be economical to set up a circuit only to use half of it, the connection from A to B, to transmit a single unit of data.

7.2.10.3 Super Services

In addition to the standard telephone services via geostationary satellites and the big LEO services, the term "super services" is used for telephone via geostationary satellites using cell phones. This super service was successfully introduced in 2000 with the geostationary Thuraya satellites, simultaneously with the bankruptcy of the big LEO systems Globalstar and Iridium. In the meantime INMARSAT also offered a super service.

7.2.10.4 Mega Services

On the scale from little (lowest bit rates) to big and super (data rates allowing voice communication), mega is next, leading to bit rates in the megabit per second range. For this service it was also proposed to use low orbiting satellites: TeleDesic (TeleDesic Inc.), Celestri (Motorola Corp.) and SkyBridge (Alcatel Espace).

TeleDesic promised the user transmission rates of up to 1.5 Gbit/s. For this purpose, 840 satellites, later reduced to "only" 288, were to fly in circular LEOs. Since bit rates of this order of magnitude require the operation of directional ground antennas with large apertures, the users would have had to operate at least two (tracking) antennas, at considerable investment and operating expense.

None of these mega LEO systems was realized. Mega services via geostationary satellites continue to be offered and operated in all communication satellite systems.

7.2.11 Transponders

Typical transmission bandwidths are 1 GHz (500 MHz in each of the two polarizations; 4.2 GHz in the case of ASTRA), not reflected in an extremely wide band, but subdivided into frequency units of, for example, 40 MHz, which are processed separately. The repeater, which converts a 40 MHz band unit into the desired frequency and transmits it amplified back to Earth, is called a transponder (it transposes the frequency it receives to the transmit frequency).

This subdividing of the actively used spectrum in transponders was necessary in the early days of satellite technology because the transmit power which would be necessary to emit the entire bandwidth was not technically available. Even today, when this satellite transmit power is technically available, transponders taking up only part of a bandwidth are used because they can flexibly serve individual antennas for given coverage zones (something that would not be possible for an *en bloc* converter). Satellites which transmit in the C-, Ku- and Ka-bands with a total of 100 transponders are state of the art.

Also state of the art are flexible payloads – flexible as to bandwidth, power and service region on Earth.

7.2.11.1 Variable Transponder Transmission Power

If one desires to change the transmit power of a transponder during flight without modifying the power output of the transmitter amplifier, the easiest method is to provide additional amplifiers: transponders which potentially have requirements for higher power are connected to additional amplifiers (parallel connection). This can be accomplished in orbit from the ground as required.

Paralleling of transmit amplifiers goes back to the first communications satellites Intelsat I and Intelsat II. The 6 W amplifiers available in 1964, which could only serve single carrier operation in Intelsat I, led to a single transponder on Intelsat II with four of these amplifiers, allowing multiple carrier operation with some six ground stations in six countries, in the USA and in Europe.

The manufacturer of these satellites continued to produce parallel connections of amplifiers in the satellites up until today – a small expense with great benefits because, should amplifiers fail on the satellite, the amplifiers intended for the parallel connection could be used to maintain operation. A transponder with only one amplifier is more useful than a transponder without an amplifier.

7.2.11.2 Multimode Tubes

The second possibility to achieve variable transponder transmit power is to construct the amplifier so that it can be operated in two or three different output stages. While any amplifier can be operated with low output power by starting with low input power, it will require for such operation the same power supply it would require if operated to full capacity.

A bi- or tri-mode amplifier can be operated at two or three output stages, in which only one-half, one-third or two-thirds of the full power supply is required.

The use of tri-mode amplifiers can be traced back to MARISAT, the first mobile communications satellites, in 1976. Starting with a slowly growing number of customers and an inefficient use of the fill factor of the satellite, the MARISAT program leased capacity to the US government. This contract included the unique demand for lessening transmission capacity

over time. In the first phase the capacity for the US government was 90% of the total, and the transponder for commercial radio service to ships and on land was kept to 10%. Later, the capacity for the government was reduced to 70% and commercial broadcasts increased to 30% etc. This achieved a smooth ramp-up of a new service for the commercial sector, unique in the history of communications satellites.

Efficient bi-mode and tri-mode traveling wave tube amplifiers are also produced in Europe and are today a component of modern satellite payloads.

7.2.11.3 Variable Transponder Bandwidth

With the use of digital bandpass filters it is possible to change the transponder bandwidth in orbit during mission operation. In the ideal situation where all 40 MHz transponders are equally used to capacity, it is not used in practice. There always are, however, transponders which are not used to capacity and others for which there is more traffic than they can transport.

With transponders whose bandwidth can be changed in orbit, the actual traffic load can be matched. No transponder is operated with unused bandwidth which picks up noise, and no traffic is rejected because a transponder is already being used to capacity.

The additional power required by a transponder whose bandwidth was increased can be supplied by transmit amplifiers connected in parallel, or by using multimode amplifiers. In the same way, power is reduced in transponders whose bandwidth was reduced in keeping with the traffic load.

7.2.11.4 Smart Antennas

Hand in hand with adjustable transponder bandwidths and transmit power go antennas which can be reconfigured in orbit (smart antennas). For a communications satellite, quite often the traffic profile changes during the mission, especially for missions lasting 15 to 20 years, a duration which is technically possible today and is used.

Even for a moderate mission life of 10 years this can be the case. One example is DFS-Kopernikus in Germany, whose antennas had been customized for coverage of the Federal Republic in the 1980s and then was unable to provide services to the formerly East German states after reunification.

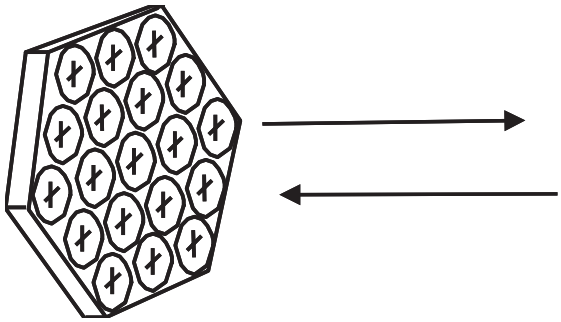


Figure 7.2.8: *The planar phased array antenna which can also be conformably shaped and simultaneously receive from and transmit to several satellites.*

Simple circular or elliptical service regions can be generated with circular or elliptical satellite antennas. Often the topological requirements are more likely to be irregular, if not geographically bizarre. To meet such requirements, suitable satellite antennas are required, and there are at least three designs available.

Quite common are passive reflectors in whose focus there is an arrangement of active feed elements which can be controlled from any service region on the ground.

Phased array antennas (Figure 7.2.8) are flown on Globalstar and Iridium, among other platforms.

Intelligent antennas are also irreplaceable on the ground, for example on vehicles: a thin-foil phased array is installed to conform to a car roof and dynamically tracks several satellites without affecting the c_w value of the vehicle. To track several satellites with parabolic antennas, several antennas would be required, and these could not be mounted on cars, much less on aircraft.

7.2.11.5 Nulling Satellite Antennas

Commercial communications satellites are increasingly subject to unintentional interference, and military communications satellites have always been protected against intentional interference. The phased array antennas (see Section 7.2.11.4), in addition to the ability to produce customized, oddly shaped footprints on the ground, can also counteract one or

several sources of interference aimed at the satellite (uplink nulling) and for that reason have for years been state of the art for military communications satellites.

7.2.11.6 Array Antennas

As in cell telephony on the ground, several individual cells are used in satellite communications instead of one conglomerate illumination for the entire coverage area. The reason is the same in both cases: the frequency bandwidth of one cell can be reused in the cell after the next one. This frequency reuse is absolutely required to accommodate continually increasing communications traffic in a physically limited, not increasing frequency spectrum.

If one-third of the entire bandwidth is allocated to each cell in a honeycomb structure and n cells are operated, then the net usable frequency band increases by a factor of $n/3$. With 200 cells on INMARSAT-4 this leads to 67 MHz of usable bandwidth per 1 MHz of spectrum.

7.2.11.7 On-board Processing and Switching

Particularly for satellite payloads serving multiple cells on the ground, the problem of switching traffic between these cells arises. To provide links within the cell as well as with any one of the $n - 1$ other cells requires $n^2/2$ connections. In conventional technologies this requirement would lead to the installation of n^2 transponders (for each connection two transponders are required, one from A to B and one from B back to A).

For example, in the case of $n = 200$ cells, this would require the installation and operation of $n^2 = 40\,000$ transponders, which is practically impossible. Instead, one employs a processor in the payload, the on-board processor (OBP), which takes over this switching function (on-board switching, OBS).

The control of the switch can take place either on-board or on the ground (the ground network controlling commands the on-board processor). In the first case, the processor must demodulate the signal in order to be able to read the desired address in its baseband. This is technically possible today and saves time and transmission capacity; the signal can be forwarded directly to the addressee, whereas with ground control the signal has to travel from the sender via the satellite to the ground station and after the switching then back via satellite to the addressee.

Demodulation of the signal in the satellite has the additional advantage that signal errors incurred on the uplink can be corrected on-board, so that the signal emitted on the downlink is "like new," free of errors.

The signal can also undergo another modulation in the satellite to optimize it for the downlink, something quite common in military spacecraft.

In addition to the dramatic advantage of the on-board processor of not having to install, say, 40 000 transponders, the operation of this processor has systems benefits.

When a ground cell in a transponder architecture has little traffic or remains altogether unused, neither its bandwidth nor its power can be transferred to another cell (another transponder) which might be overloaded and must reject additional traffic.

By contrast, the on-board switching handles all traffic, no matter how heterogeneously distributed over the cells, flexibly assigning bandwidth and power to cells needing it, in real time and without rejecting any traffic.

7.2.11.8 Payloads with Multiple Frequency Bands

Of the frequency bands listed in Section 6.4.2 allocated to satellite communications, the P-, L, S-, C-, X-, Ku-, Ka-, Q-, V- and W-bands, current national communications satellites commonly use four, namely the C-, Ku- and Ka-bands, in commercial communication: the Ku-band for broadcasting, the X- and Ka-bands for military telecommunications, and the L- and S-bands for mobile communications.

Cross-connections between these bands (cross-strappings) in the payload usually assure technical flexibility. At least the P-band with its limited bandwidth will be cross-connected, to link the anchor station in the Ku- or Ka-band with the satellite (feeder link) and operate with the user in the P-band, with its narrow range of voice channels.

7.2.12 Transmission Techniques

Satellite communications use the same modulation procedures as do terrestrial communications, preferably phase shift keying modulation with 4 (QPSK) and 16 (APSK) phase states.

The QPSK modulation ($M = 4$ with 2 bits/symbol) requires 0.8 Hz/bit of bandwidth: that is, 1 Mbyte/s requires 0.8 MHz bandwidth; the 16APSK procedure ($M = 16$ with 4 bits/symbol) requires 0.4 MHz per 1 Mbyte/s, including rate 3/4 trellis coding and a guard band between carriers. The state of the art is that modulation equipment adjusts the degree of modulation to the quality of the channel in the course of transmission (DVB-S2 Standard, Self-Adjusting Modulation Techniques). As long as the transmission channel maintains nominal quality, the modems modulate 16APSK; if channel quality breaks, for example because of the weather, the modem reduces the degree of modulation, such as to QPSK. This makes the transmission twice as robust, although only half as fast, but it is better to have error-free transmission than rapid but faulty transmission.

Also for channel coding, equipment with adaptive code rates is used today in satellite communications, as it has been used terrestrially since the 1970s. If the channel degrades and the bit error rate increases, the coder/decoder reduces the code rate from, for example, 3/4 (every fourth bit is a code bit) to 1/2 (every second bit is a code bit). In this way the error rate is reduced at the cost of a slower transmission.

The third way to maintain transmission despite difficult channel conditions, next to reducing the modulation and increasing the coding, is to reduce the carrier load in the satellite transponder (load dumping). If the situation requires it, a few carriers will be temporarily offloaded and their power assigned to the remaining carriers in the transponder so that they, now equipped with higher transmit power, have a better chance of maintaining nominal quality (threat adaptive offloading).

One could also use the bandwidth freed by offloading for increased channel coding of the remaining carriers, thus making them even more robust, and thereby achieve additional error protection.

7.2.13 Multiple Access Techniques

How can many users gain access to a single resource, the communications satellite? An obvious answer is: one after the other (successive multiple access, called time division multiple access (TDMA)). But users could also work with different frequencies at the same time, in the same way that radio stations transmit on their own individual frequencies (multiple access by subdividing a frequency range, called frequency division multiple access (FDMA)). A third option is to individually code multiple transmissions so that they can be superimposed on the same frequency (multiple access via coding, called code division multiple access (CDMA)). Knowing the selected code, the recipient can filter out, from the sum of all signals, the one meant for him or her. CDMA is costly in terms of bandwidth.

TDMA is currently the method of choice by far; it achieves the highest satellite throughput for a given frequency band and power.

7.2.14 Frequency Ranges in the Electromagnetic Spectrum

The frequency spectrum officially allocated to satellite communications in the range from 7 kHz to 1000 GHz is increasingly utilized, presently up to some 100 GHz. Besides the technology available at such frequencies, propagation in the atmosphere is a limiting factor for the highest frequency.

It is interesting to note that propagation in the atmosphere does not become increasingly difficult as frequencies exceed 100 GHz. Attenuation by precipitation more or less stagnates there, so that one can also work in the infrared and visible ranges. The great advantage there is that laser links enable much larger bandwidths, so a far greater volume of data per unit time can be transmitted. And for the satellite technology it is especially advantageous that laser links require considerably less electrical power per transferred bit.

These giga-size transmissions at laser frequencies are possible with relatively small antennas (they are

called telescopes here), another systems advantage. However, these small telescopes have an extremely sharp focus – they hit a point on the Earth with an aperture of 1/1000°, which translates to a diameter on the ground of about 300 m, not even the area of an urban district.

There are two ways to reach users distributed throughout an urban district or across even larger regions: either with **terrestrial networks**, in which the gigabit link is received by one anchor station and forwarded terrestrially to the target address, or by using a **satellite phased array antenna**, which can receive a large number of laser links from distributed stations and then retransmit them. The first phased array antennas in the laser frequency range have already appeared. They have to form a sufficiently large number of beams to be able to serve a large user group, since each laser link typically serves only one user. The first large-broadband laser links (at wavelengths of 1.55 μm) from GEO to Earth have been in operation for several years, and from LEO to Earth for 30 years. The further development of this technology for satellite communication will progress rapidly. Major steps have been taken in just the last few years in the direction of:

- Significantly more powerful emitters
- Coherent modulation
- Capability to maintain operations while looking into the Sun.

In the USA 40 Gbit/s transmissions have successfully been demonstrated (1 Gbit is the entire contents of the *Encyclopedia Britannica* including tables and figures). Even a small number of such links correspond to a bandwidth which in the electromagnetic spectrum between 1 and 100 GHz is not physically available.

7.2.15 Disposal of Satellites

A severe problem associated with space activities is disposing of satellites at the end of their service life. This is significant because satellites are becoming so commonplace. Whereas LEO satellites are slowed down over time by the residual atmosphere that exists up to about 500 km altitude, and then eventually burn up as they enter Earth's atmosphere proper, ICO and GEO satellites remain in orbit for a long time at the end of their service life, where they not only block an orbit position for their successors, but in the worst case could collide with other satellites.

The frequent practice of raising GEO satellites to a "graveyard orbit" higher than GEO has for a long time not been considered a viable solution, because more and more satellites are colliding also in this orbit, and half of the ensuing debris is hitting satellites in GEO. Returning satellites to Earth after their mission life will have to become standard procedure; the sooner, the better.

7.2.16 Outlook

Over the last 40 years, satellite communications have been the money-earning workhorse of space activities, and this will surely continue to be the case for at least the next 40 years. Wernher von Braun once remarked that if all other space activities were to receive a budget amounting to some 5% of the turnover of satellite communications, then space activities would continue to prosper; this hope has so far been fulfilled.

Fixed satellite communication will continue to be irreplaceable wherever broadband transmissions have to be established rapidly and flexibly, including disassembly after a period of time. Also, **thin route traffic** will not be replaced by optical fiber technology for a long time.

Radio and television broadcasting will for a long time into the future be more economical via satellite than via terrestrial transmitters or cable networks. Terrestrial transmitters in Germany alone cost over DM200 million per year in 1985, more than the cost of a satellite which can broadcast 100 television programs for 15 years. And the satellite can serve all of Europe, certainly 100% of Germany, something that the terrestrial transmitters cannot accomplish with their coverage of less than 95% of the country.

Mobile telecommunications over large areas for surface vehicles, ships and aircraft, reliably and with adequate bit rates, cannot be accomplished terrestrially; nobody will ever lay a fiber optic cable of any bit rate to an aircraft.

For this purpose the first satellites with 40 m × 40 m antennas (ET-5) have been put into orbit; fast

40 Gbyte/s processors are being flown for on-board processing and switching. Ever since Intelsat 8 (1997), satellite lifetimes of 20 years have been realized (which is worth comparing to a motor vehicle remaining fully functional for 20 years without servicing or repair) with satellite masses up to 10 t and with 20 kW of electrical primary power still available at the end of the mission.

It has long been known how to co-locate up to 10 satellites and thereby broadcast, for example, up to 100 television programs from one orbital position, for reception by small, pizza-sized home receivers. But these satellites also transmit voice and data at a speed of megabytes per second from the geostationary orbit directly to the user's hand-held phone, and accomplish this feature over wide areas, in regions where it would not be economical to use terrestrial broadcasting towers, and also over the oceans [7.2.8].

Bibliography

[7.2.1] Morgan, W.L. *Communications Satellite Handbook*. New York: Wiley Interscience, 1998.

[7.2.2] *Via Satellite*, monthly journal, Phillips Business Information, Inc., Rockville, MD, 2007.

[7.2.3] Pritchard, W., Sciulli, J. *Satellite Communication Systems Engineering*. Englewood Cliffs, NJ: Prentice Hall, 1996.

[7.2.4] Lutz, E., Werner, M., Jahn, A. *Satellite Systems for Personal and Broadband Communications*. Berlin: Springer Verlag, 2000.

[7.2.5] Dodel, H., Häupler, D. *Satellitennavigation – Galileo, GPS, GLONASS, Integrierte Verfahren*. Heidelberg: Hüthig Verlag, 2004.

[7.2.6] *TeleSatellit*, bimonthly journal, Munich, 2007.

[7.2.7] Dodel, H., Eberle, S. Satellitenkommunikation. Ständiger Lehrgang DK2.08 der Carl Cranz Gesellschaft für technisch-wissenschaftliche Weiterbildung, Oberpfaffenhofen, 2007.

[7.2.8] Dodel, H. Eberle, S. *Satellitenkommunikation*. Berlin: Springer Verlag, 2007.

7.3 Navigation

Stefan Sassen

Satellite navigation provides precise data for coordinates in a global reference frame and as quasi-by-product data for precision time measurements. The continuous availability of this information enables new services and applications to significantly contribute toward economic growth, for example by increasing the efficiency of transport and logistics. The fundamentals of satellite navigation and future developments are outlined in this section.

7.3.1 Basic Principles of Satellite Navigation

7.3.1.1 History of Navigation

In 1519 when Magellan set sail on his journey around the world, he was equipped with sea charts, globes, wooden and metallic theodolites, quadrants, compasses, magnetic needles, timepieces and "a log towed astern." With these tools and with many years of experience it was possible to determine the speed of the vessel, its direction and latitude [7.3.1]. An accurate **determination of longitude** during voyages off-shore was made possible only about 250 years later by the development of sufficiently accurate clocks. For example, during a two-week voyage, using a clock with a time accuracy of 1 PPM would result in a time error of more than 1 second, which at locations near the equator causes a positional error of more than 500 meters. In 1761 John Harrison received a prize offered by the British Parliament for the off-shore determination of longitude. This prize was awarded for his pendulum clock No. 4, which kept time accurately within only 5 seconds during an 81-day voyage. The astronomical determination of longitude based on time measurement using a clock on-board the vessel became obsolete only through the development of telegraphy and radio technology at the end of the nineteenth century. During the period from the fifteenth up to the beginning of the nineteenth century, navigational requirements provided the

impetus for **clock technology** developments. Today global navigation satellite system (GNSS) technology is very frequently used for time synchronization over large distances, for example, for telecommunications purposes.

Since antiquity sailors near the coast had used beacons[1] and charts on which their positions were registered. The identification of the individual beacons was made from their periodically coded light signals. Determining position at sea resulted from **triangulation**. Encoding and accurate knowledge of the location of signal sources as well as triangulation and/or trilateration are also the basis for determining location by means of the long-range navigation (LORAN) system and GNSS. The encoding of the transmitters corresponds to the CDMA (Code Division Multiple Access) technology used today for distinguishing different GPS satellites.

Three technological disciplines have determined navigation methodology over the course of the centuries:

- Geodesy and charts
- Time measurement
- Astronomy, and today space technology.

Geodesy and Cartography: The basis for any type of navigation is utilization of a uniform reference system and representation of the Earth's surface in charts. In the course of the nineteenth century the increasing accuracy of geodetic measurements revealed that a representation of the world as a sphere or as an ellipsoid is not sufficiently accurate. With the scientific discovery of continental drift and the mobility of the Earth's surface, which could be measured for the first time in the 1950s, the assumption of a rigid Earth surface also had to be given up.

Time Measurement: The development of highly accurate atomic clocks with a time stability error of less than 2 ns per day was indispensable for measuring signal propagation time with sufficient accuracy. Today cesium or rubidium atomic clocks are used on-board navigation satellites. For the Galileo system it is planned to use even more accurate hydrogen masers.

[1] The Colossus of Rhodes (~280BC) and The Pharos of Alexandria (~250BC) are well-known beacons of antiquity. Today lighthouses are still in operation to warn of dangerous locations and/or coastal regions or in cases when other navigational means are faulty or unavailable.

Space Technology: In order to be able to transmit the timing signals from an accurately known position in orbit, the orbital motion of the satellite must be highly stable and predictable. Decades of experience in space flight technology allow prediction several days in advance with an accuracy of a few meters in the orbital motion of navigation satellites, which have an orbital speed of 3.9 km/s.

7.3.1.2 Origin of Satellite Navigation – TRANSIT

TRANSIT (also known as NAVSAT) was the first satellite navigation system and the predecessor of GPS. Its transmission frequency was tuned to be 150 and 400 MHz. Starting from 1958, TRANSIT was developed by the US Navy. It was put into service for the first time in 1964 and after 1967 it was also used for civilian purposes [7.3.2]. The TRANSIT satellite system was decommissioned on December 31, 1996.

TRANSIT was the first-ever functioning satellite navigation system. It was originally developed for commanding the ballistic missiles of the US Navy. An accurate missile guidance system was urgently requested at that time, mainly for the nuclear weapons on submarines and aircraft carriers.

The achievable **accuracy** depended upon receiver technology and amounted to between 500 and 15 meters. Under normal operating conditions, the TRANSIT system required at least four functional satellites. In the final phase of TRANSIT operation the constellation consisted of six satellites: three operational satellites and three used as a backup system. The TRANSIT satellites used polar orbits with an altitude of approximately 1220 km.

The use of four ground stations, which transmitted correction data every 12 hours to the satellites, was necessary. The TRANSIT system was controlled by the Naval Astronautics Group (NAG) in Point Mugu, California.

7.3.1.3 Functional Principle of Satellite Navigation

Today navigation with satellite signals is performed with a so-called **trilateration procedure**, which allows the calculation of the unknown three-dimensional coordinate by measuring distances to transmitters at

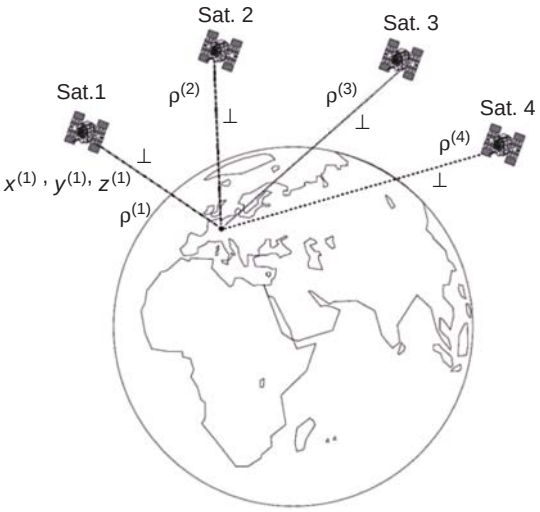


Figure 7.3.1: *Principle behind satellite navigation to determine position with four satellite signals.*

well-known locations [7.3.3], [7.3.4]. The transmitters are on-board satellites with precisely determined orbits which are transmitted to the users by the navigation message. The determination of the distance between user and satellites is made by measuring the signal propagation time from the satellite to the user. The satellite and receiver clocks must be accurately synchronized for the propagation time measurement since the signal propagates at the speed of light, about 0.3 m per 1 ns. On-board the satellites are highly accurate atomic clocks whose time variances are closely monitored and modeled to achieve the required accuracy. Since such precise clocks would be very expensive and unpractical in users' receivers, the receiver time is synchronized by the signal of a fourth satellite (Figure 7.3.1).

For this reason at least four satellite signals are necessary to determine a three-dimensional coordinate (x, y, z) and the local time t. The signals are transmitted from the satellite's position $x^{(k)}$, $y^{(k)}$, $z^{(k)}$. Then, by taking the receiver clock bias b into account, it follows for the **pseudo-range measurements** $\rho^{(k)}$ that

$$\rho^{(k)} = \sqrt{(x^{(k)} - x)^2 + (y^{(k)} - y)^2 + (z^{(k)} - z)^2} - b \qquad (7.3.1)$$

with $k = 1, 2, \ldots, N$.

7.3.2 Satellite Navigation Systems

7.3.2.1 System Design

A complete satellite navigation system comprises the following segments: space segment, ground segment, transfer segment and user segment.

The following are basic considerations for the design of all satellite navigation systems:

- A **passive system**, which allows the user to determine a position with only received signals and without sending any signals (broadcast system). Such a system provides capacity for any number of users without the need to adapt the system. In addition, it is desirable for military applications that the user does not radiate any signal which could reveal his or her position. Due to progress in clock technology, around 1970 the decision was made that satellites should broadcast a **time synchronous signal** and that the **trilateration** principle be used instead of a Doppler or hyperbolic positioning method.
- Due to the general progress in signal processing technology, continuous transmission of **CDMA** signals was chosen for GPS, since it was considered to be superior to pulsed or FDMA (Frequency Division Multiple Access) signals.
- For the **carrier frequencies** the choice fell on L-band frequencies (1–2 GHz) because of their relative insensitivity to tropospheric weather influences, the moderate propagation dependence on ionospheric conditions and the availability of a sufficiently large frequency spectrum (approx. 20 MHz per signal) in the 1970s.

For the design of the segments, the following basic considerations were taken into account.

Space Segment

Navigation satellites send their signals to Earth from accurately known orbital positions. The core elements of each navigation satellite are highly accurate **atomic clocks**, from which the highly stable carrier frequency is derived, and **signal generators**, which modulate the carrier frequency with the navigation message and the time signal.

The space segment consists of a constellation of satellites, so that users have as frequently as possible at least four satellites in direct line of sight. The choice

of orbits for a navigation satellite constellation is determined by the following considerations.

LEO (Low Earth Orbit) constellations with an orbital altitude < 2000 km provide satellite visibility of less than 20 minutes and due to the large orbital velocity the signals show a strong Doppler shift. For global coverage with at least four visible navigation satellites, a constellation with more than 100 satellites would be required. The advantages of a LEO constellation would be the moderate launch costs per satellite, and the fact that the signals could be broadcast with relatively low power density. On the other hand, orbital disturbances caused by atmospheric friction would be large.

MEO (Medium Earth Orbit) satellites with orbital altitudes around 20 000 km and a period of approx. 12 hours have a visibility of several hours for users on the Earth's surface, which significantly reduces the number of signal reacquisitions compared to LEO satellite signals. The launch costs are higher, but 24–36 satellites per constellation are sufficient to achieve worldwide navigation coverage. The satellite orbits are more stable, but cosmic radiation affecting satellite components is significantly higher, which limits in particular the life cycle of key components (atomic clock, signal generator).

GEO (Geostationary Orbit) satellites with an orbital altitude of approx. 36 000 km remain at a fixed horizontal position and for worldwide coverage only few satellites are required. But due to the equatorial position of the GEO satellites, the geometrical distribution would deteriorate the achievable navigation accuracy and at higher latitudes satellite visibility would be easily blocked by buildings, trees, etc.

For today's navigation constellations (GPS and GLONASS) and also for Galileo, MEO satellite orbits were selected. For augmentation systems (e.g., WAAS, EGNOS, MSAS) GEO satellites are preferred [7.3.5]. A combination of GEO and MEO satellites is planned for the Chinese COMPASS system.

Furthermore, Japan is planning to launch a mini-constellation consisting of three satellites in HEO orbits (highly inclined elliptical orbits) to be used as an augmentation system.

Ground Segment

The ground segment comprises all terrestrial components which are required to operate the complete navigation system and to ensure its reliability and the quality of the broadcast signals. This includes the control centers (ground control center (GCC)) as well as a number of remote sites and an associated communication network between the remote sites and the control centers.

The ground segment is divided into a ground control segment (GCS) and a ground mission segment (GMS).

The **GCS** is used to command the satellite fleet for both routine and special operations, such as commissioning the satellites (launch and early operation (LEOP) and in-orbit testing (IOT)). The command and control signals are generated by the GCS equipment in the GCC and are transmitted over a dedicated communications network to the telemetry, tracking and control (TT&C) stations, which send the signals to the navigation satellites and also receive telemetry data from them.

The **GMS** is responsible for the correct generation of timing and navigation signals. For this purpose the signals which are broadcast from the satellites are continuously monitored with a global network of sensor stations and passed on to the control center. In the GCC the respective control commands and integrity data are generated and transmitted over a communication network to uplink stations (ULS) sending the GMS command to the satellites.

Transfer Segment

Taking Galileo as an example, the selected launcher will transfer the satellites of the constellation directly into MEO, using double or multiple launchers, depending on the capacity of the selected rocket.

User Segment

The user segment refers to the **navigation terminal** (or radio) consisting of antenna, receiver electronics and output unit for the position data. Today a large variety of terminals are available for different accuracy classes (centimeters to meters), forms and sizes (ranging from shoebox-sized terminals for precise geodetic applications down to single-chip receivers which are built into mobile telephones). The cost of a simple receiver chip is below €10 today, whereas very precise two-frequency receivers still cost over €10 000.

7.3.2.2 GPS, GLONASS, Galileo and COMPASS

The first satellite navigation system available worldwide was **GPS**. On February 21, 1978 the first GPS satellite was launched into orbit on-board a Delta IV rocket. Although GPS was used intensively for military operations during the Gulf War in 1991, the system did not reach full operational capability before 1995.

GPS is operated and further developed exclusively by the US military. It is under the control of the US Space Command (USSPACECOM) in Colorado, which is responsible for commanding and controlling all US military space activities. Even if civilian GPS applications come to predominate and large worldwide commercial interest exist in GNSS applications, the whole GPS system itself remains a military system.

During the Cold War, the Soviet Union advanced the development of its own military navigation system, named **GLONASS**. However, due to the limited lifetime of the first GLONASS satellites and the lack of funds at the end of the Cold War, full operational capability has not yet been achieved. In recent times increasing efforts have again been undertaken to further develop and complete the GLONASS constellation.

The most important difference between GPS and GLONASS is the signal type. All GPS satellites transmit on the same frequencies. The distinction among individual satellites is achieved by the CDMA signal structure (which is also used for mobile phone signals), while GLONASS satellites are identified by slightly different transmitter frequencies (FDMA principle with 8 MHz separation). This prevents the use of signals from both systems at the same time with simple cost-efficient receivers.

Europe is in the process of developing and deploying the first civilian navigation system, called **Galileo**. At present the first two test satellites are in orbit and the first four satellites of the constellation are under development. The completion of Galileo is planned for the year 2013.

Additionally, China has advanced its plans to develop its own system, named **COMPASS**. Respective registrations for frequencies and orbital positions have been filed and the first test satellites have been deployed in orbit. At present the system comprises five geostationary satellites, three satellites in inclined geosynchronous orbits and 27 satellites with MEO orbits at an altitude of 27 840 km. The geostationary satellites are called Beidou, the Chinese name for the astronomical constellation Ursa Major. Beidou-1A and 1B are test satellites that were launched on October 30 and December 20, 2000. The next generation of satellites will become part of the Chinese navigation system. They are planned to take positions over the Asia–Pacific area at 58.75°E, 80°E, 110.5°E and 140°E. Beidou-2A (formerly Beidou-1C) was launched on May 24, 2003. The fourth satellite was launched into orbit at the beginning of February 2007. The first MEO satellite was launched in April 2007.

Since China is also involved in the European Galileo system at the same time, it remains unclear how COMPASS will evolve in future.

Table 7.3.1 gives an overview of the characteristics of the different GNSS systems [7.3.3], [7.3.4], [7.3.6].

7.3.3 Space Segment

7.3.3.1 Navigation Satellite Components

Navigation satellites consist of standard satellite subsystems (see Chapter 4) and specific payload subsystems dedicated to navigation functions (e.g., atomic clocks, signal generators). The most important components are described below using the Galileo satellite as an example:

On-board the Galileo satellites two passive hydrogen masers and two rubidium atomic clocks are planned. Each is used to generate the frequency and time standard to high accuracy, stability and the lowest possible noise level.

For frequency generation using **passive hydrogen masers** the hyperfine structure transition ($\Delta E \approx 10^{-5}$ eV or $f \approx 1.420$ GHz) of atomic hydrogen is used. To this end the activated hydrogen atoms are fed into a quartz resonator, in which the atoms radiate at highly stable frequencies due to the long decay time of about 1 second. This frequency is amplified in a resonant mode by external electronics and used as the frequency standard for signal generation on-board the satellite. The resulting time stability is better than 0.5 ns in 12 h.

Table 7.3.1: *Characteristics of various satellite navigation systems.*

	GPS	GLONASS	Galileo	COMPASS
Nominal constellation	MEO (21+3/6) incl. 3 active satellites	Walker MEO (24/3/2)	Walker MEO (27/3/1) + 3 inactive satellites	GEO (5) + IGSO (3) + MEO (27)
Actual constellation (Jan 2009)	31 satellites, distribution 6–5–5–5–5–5	16 active and 1 inactive satellites	2 test satellites (GIOVE A&B)	Beidou-1A, 1B, 2A, 2B (GEO-Sat) Compass M1 (MEO)
MEO orbit	Circular, 55° inclination	Circular, 64.8° inclination	Circular, 56° inclination	Circular, 55° inclination
Half-axis	26 560 km	25 440 km	29 601.3 km	27 840 km
Orbital period	11 h, 58 min	11 h, 15 min	14 h, 4 min	
Mass	Block II: approx. 1660 kg Block IIA: approx. 1816 kg Block IIR: approx. 2023 kg Block IIF: approx. 1545 kg	GLONASS: approx. 1415 kg GLONASS-M: approx. 1500 kg GLONASS-K: approx. 750 kg	1 to 4, approx. 680 kg	
Clock type	Rubidium, cesium	Cesium	Rubidium, H-maser	
Georeference system	WGS84 based on ITRF	PZ-90	GTRF based on ITRF	
Time reference	UTC (NO)	UTC (Russia)	UTC (GTSP)	
Modulation	CDMA	FDMA	CDMA	CDMA
Ground segment	1 master control station 6 monitor stations 3 antenna sites	1 system control centre 5 TT&C stations	2 ground control centres 5 TT&C sites 9 ULS sites about 40 sensor sites (GSS)	

Up-to-date information on the status of the GPS satellites can be found at http://www.navcen.uscg.gov/gps/default.htm.

7

Rubidium atomic clocks are the second type of clock on-board a Galileo satellite. This type is also used on-board GPS and GLONASS satellites. A time stability of approximately 1.5 ns in 12 h is achieved by using the hyperfine structure line of rubidium-87 at 6.835 GHz. If rubidium atoms in a hot gas are pumped with an external microwave at exactly this frequency, then the rubidium atoms are activated from their ground state to a higher energy state in which the absorption of an external rubidium light source reaches its maximum. A photodiode determines this absorption peak, and the frequency of the external microwave is adjusted to keep the absorption peak at maximum.

For monitoring and controlling the four clocks of a Galileo satellite, the so-called **clock monitoring and control unit** is used, which also interfaces with the signal generator. For example, it must be always ensured that the master clock and the active redundant clocks are synchronized in order to ensure seamless switching in case of master clock failure.

The **navigation signal generator unit** (NSGU) combined with the frequency upconverter unit generate the navigation signals (carrier frequency, modulation and navigation message) based on the master atomic clock frequency standard and the data stream received from the GCC.

These L-band signals on-board the Galileo satellites are then amplified by **solid-state power amplifiers** (SSPAs). At a transmission power level of 50 W per signal, a minimum dissipation loss and a maximum of linearity are required.

The antenna system consists of an L-band antenna for transmitting the navigation signals in the

frequency range of 1200–1600 MHz. The C-band antenna receives mission data from the uplink stations, such as clock parameters, orbit parameters and integrity data. The S-band antenna is used for sending and receiving the telemetry (TM) and telecommand (TC) data. Additionally, the S-band channel serves as backup for ranging (at meter accuracy) and as backup channel for the uplink of mission data.

The laser retroreflector is used for measurements of satellite orbit data to subdecimeter accuracy. These measurements are not part of the routine operations procedures but used for monitoring and calibration. Only a few GPS and some GLONASS satellites are equipped with laser reflectors.

To protect the TM/TC data as well as to encode the protected signals (CS and PRS) special **cryptographic components** (payload and platform security unit) are used on-board the Galileo satellites.

7.3.3.2 Satellite Orbits

The prerequisite for determining receiver position by means of trilateration is the continuous and exact knowledge of **transmitter position**, that is the accurate orbital position of the transmitting antenna center point. Using the orbit parameters transmitted in the navigation data message, the position can be predicted with an accuracy of a few meters. This is even more remarkable if one considers that the orbital altitude of the satellites is 26 560 km for GPS and 23 616 km for Galileo and that the prediction is based on measurements which are usually older than 10 hours.

Ideal Orbits

Kepler's laws describe to a first approximation the movement of navigation satellites around the Earth [7.3.7]. These can be derived from Newton's laws.

For a central gravitation potential and the respective equation of motion the following applies:

$$\boldsymbol{V}(r, \boldsymbol{\Phi}, \boldsymbol{\lambda}) = \frac{GM}{r}, \quad \ddot{\boldsymbol{r}} = -\frac{GM}{r^3}\boldsymbol{r} \tag{7.3.2}$$

Solving this equation of motion results in an elliptical orbit, with the Earth's center of gravity at one of the two focal points. The satellite orbits are preferably described by the so-called six **Keplerian elements** (see Section 2.2).

Perturbation of Ideal Orbit

Deviations from the spherical gravitational potential and additional force excited by other causes have an extremely small impact on the satellite orbits, but to achieve their computation with a precision of a few meters or better, these effects must be taken into account.

Noncentric Gravitational Forces of the Earth: The Earth is not absolutely spherical; its density distribution is not homogeneous (or spherical) and this distribution varies over time. To the next order of approximation the Earth is described as a spheroid with its equatorial radius being about 20 km larger than the polar radius. The **gravitational potential** at distance r from the Earth's center, at geocentric longitude λ and at geocentric latitude Φ is described by the following spherically harmonic series expansion [7.3.9]:

$$V(r, \Phi, \lambda) = \frac{GM}{r}\left[1 + \sum_{n=2}^{n_{nm}} \sum_{m=0}^{n} \left(\frac{a}{r}\right)^n \cdot \bar{P}_{nm}(\sin \Phi) \times (\bar{C}_{nm} \cos m\,\lambda + \bar{S}_{nm} \sin m\,\lambda)\right] \tag{7.3.3}$$

where:

a = semi-major axis of the WGS84 ellipsoid (World Geodetic System 1984, a three-dimensional global coordinate system; all details are given in [7.3.9]),

$\bar{C}_{nm}, \bar{S}_{nm}$ = normalized gravitational coefficients,

$\bar{P}_{nm}$ = normalized associated Legendre functions.

The substantial influence of higher order terms in comparison to the spherical central potential follows from the Earth's flattening at the poles and the larger expansion at the equator.

Gravitational Forces of the Moon and the Sun: Beyond the Earth's gravitational forces, the gravitational forces of the Moon and the Sun also affect the satellite orbits, with the Moon having the larger effect (just as in its influence on the generation of tides). Tides on the Earth's surface caused by the Moon and the Sun and the associated variations in mass distribution on the Earth affect satellite orbits to a smaller extent and are usually neglected in standard models predicting navigation satellite orbits.

Table 7.3.2: *Forces affecting satellites.*

Force	Acceleration [m/s^2]
Central gravitational force	0.56
Higher order terms of Earth's gravitation	$5 \cdot 10^{-5}$
Lunar and solar gravitational forces	$5 \cdot 10^{-6}$
Solar radiation	10^{-7}

Forces of Solar Radiation: The radiation pressure caused by the particles and the photons emitted by the Sun also affect navigation satellite orbits. The effect depends on the satellite's attitude to the Sun and on the Earth's shadowing effect.
Considering these perturbation forces, the **equation of motion for the satellites** is modified as follows:

$$\ddot{\boldsymbol{r}} = -\frac{GM}{r^3}\boldsymbol{r} + \frac{1}{m}\boldsymbol{F}(\boldsymbol{r}, \dot{\boldsymbol{r}}, t) \qquad (7.3.4)$$

In summary, the forces and the resulting accelerations which affect satellite motion are shown in Table 7.3.2 in descending order of magnitude.

These effects must be taken into account for any position determination based on satellite navigation signals. Therefore the GPS data message contains an extended set of parameters (so-called quasi-Keplerian elements) describing the satellite orbits.

GPS Orbits

The GPS satellites have an orbital velocity of 3.9 km/s and a period of half a sidereal day, which is about 11 h 58 min. The mean distance to the Earth's center is 26 560 km. The nominal GPS constellation consists of six planes with each plane having four equally spaced satellite positions, although today more than 24 satellites are usually active. The six orbital planes all have an inclination of 55° relative to the equatorial plane. The right ascensions of the ascending node for the six orbital planes are separated by 60° in the equatorial plane. The 55° inclination results in satellite positions between 55°N and 55°S latitude. The satellites of block I had a different orbital plane inclination of 63°.

With this GPS constellation, the polar regions have below-average coverage (in contrast to the TRANSIT system, which had above-average coverage of polar regions due to its polar orbits). As a result, signal reception at high northern and southern latitudes is only possible with accordingly lower elevations. On the other hand, this GPS constellation shows very good and homogeneous satellite coverage for the more densely populated areas between these latitudes, which guarantees high availability with at least four satellites in view. Additionally this design leads to a relatively stable constellation since interfering factors (e.g., gravitational fields, solar winds) on average affect all satellites in the same way.

Ephemeris

The computed satellite orbits and velocity data are called satellite ephemerides. The GPS ephemerides are computed with a so-called **quasi-Keplerian model**, which has an extended set of 15 (instead of 6) Keplerian elements. Parameter No. 16 shows the reference time for the set of ephemeris parameters, and parameter No. 17 (issue of data ephemeris (IODE)) indicates any alteration to the set of parameters. The following list from the GPS Interface Specifications (IS-GPS-200D, 2004) shows all parameters (Table 7.3.3).

The three parameters Δn, $\dot{\Omega}$ and IDOT describe the linear variation over time of the mean anomaly, the angle of the ascending node and the inclination angle.

Three pairs of parameters (C_{xs}, C_{xc} with $x = u, r, i$) describe the second-harmonic term of the satellite orbits as a function of the argument of latitude ϕ (angle of the satellite position in the orbital plane measured from the equator). One pair of parameters each (C_{xs}, C_{xc}) describes the correction of the argument of latitude, the correction of the orbital radius and the correction of the inclination angle in accordance with

$$C_{xc}\cos(2\Phi) + C_{xs}\sin(2\Phi) \qquad (7.3.5)$$

Every two hours a new set of the quasi-Keplerian elements is used for transmission in the navigation message. The quasi-Keplerian elements are generated in the master control station (MCS), and with each uplink sequence to a GPS satellite a set of parameters for the next 14 days is sent. The parameters computed in advance would result in decreasing accuracy of the orbital estimations with increasing latency between

Table 7.3.3: *List of parameters transmitted in the GPS navigation message.*

Symbol	Explanation
M_0	Mean anomaly at reference time
Δn	Mean motion difference from computed value
e	Eccentricity
$\sqrt{A}$	Square root of the semi-major axis
Ω_0	Longitude of ascending node of orbit plane at weekly epoch
i_0	Inclination angle at reference time
ω	Argument of perigee
$\dot{\Omega}$	Rate of right ascension
IDOT	Rate of inclination angle
C_{uc}	Amplitude of the cosine harmonic correction term to the argument of latitude
C_{us}	Amplitude of the sine harmonic correction term to the argument of latitude
C_{rc}	Amplitude of the cosine harmonic correction term to the orbit radius
C_{rs}	Amplitude of the sine harmonic correction term to the orbit radius
C_{ic}	Amplitude of the cosine harmonic correction term to the angle of inclination
C_{is}	Amplitude of the sine harmonic correction term to the angle of inclination
t_{oe}	Reference time ephemeris
IODE	Issue of data (ephemeris)

the calculated point in time and their use in the navigation message. Therefore, the transmission of the set of parameters takes place usually once per day for each GPS satellite. The parameter sets stored on the satellite serve for emergencies in which no daily transmission is possible.

Almanac

Each GPS satellite constantly transmits its complete quasi-Keplerian elements. Additionally, each GPS satellite transmits a parameter set for the **complete GPS constellation** with reduced accuracy in the form of an almanac. This almanac is helpful for receivers acquiring signals; for example, the signal search area can therefore be limited to the number of satellites theoretically visible over the horizon, or when the ascent of a satellite over the horizon is expected, the signal search can be started. The almanac parameters provide only for a linear clock model and the set of quasi-Keplerian elements comprises only seven parameters.

Precise Ephemeris

Precise ephemerides can be made available by **terrestrial services**. For example, the IGS (International GNSS Service) has a network of over 350 reference stations worldwide. At present the following data is available from the IGS: **ultra rapid predicted** (ephemeris in real time with approx. 10 cm accuracy), **ultra rapid observed** (three hours of latency and accuracy < 5 cm), **rapid** (17-hour latency and accuracy < 5 cm) and **final** (13-day latency and accuracy < 5 cm).

Galileo Orbit

The Galileo constellation is planned as a Walker (27/3/1) constellation in order to achieve the maximum possible complete and continuous coverage of the Earth's surface.

A **Walker constellation** ($T/P/F$) consists of satellites in circular orbits with equal orbital altitude H and inclination I. A Walker constellation is characterized by the following set of parameters: orbital altitude H, inclination I, total number of satellites T, number of orbital planes P (whose nodes are equally distributed at the equator) and relative distance F between satellites in neighboring planes ($F = 0$ to $P - 1$ in multiples of the pattern unit $PU = 360°/T$).

For the Galileo constellation the semi-major axis $A = 29\,994$ km, which corresponds to a mean orbital altitude of $H = 23\,616$ km. The Galileo orbits have an inclination of $I = 56°$.

The Galileo constellation will consist of $T = 27$ satellites, which are placed on $P = 3$ orbital planes. Thus the pattern unit $PU = 360°/27 = 13\,1/3\,°$. This results in a distance of the satellites in one plane of $360° \cdot P/T = PU \cdot P = 40°$ and a distance between nodes of $360°/P = 120°$. The phase difference between the Galileo satellites in successive planes is $PU \cdot F = 13\,1/3°$.

Additionally, a spare satellite will be placed in each orbital plane, so that 30 Galileo satellites will be deployed altogether on three orbital planes.

The sidereal period of a Galileo satellite amounts to 14 h 4 min 42 s and after 17 revolutions, or 10 days, the satellites again reach their starting points.

7.3.4 Ground Segment

The ground segment comprises all installations on the ground which ensure the operation of the overall system and which guarantee the quality and performance of the transmitted navigation information. This primarily concerns ground control centers and ground stations. In the case of Galileo two GCCs are planned in Germany and Italy, forming the backbone of the Galileo ground segment.

The Galileo ground segment consists of two major parts: the ground control segment (GCS) and the ground mission segment (GMS).

The **GCS** is responsible for the satellites' platform operations and their correct orbital positions. Five TT&C stations are available, which are globally distributed and which transmit telecommands and receive satellite monitoring data utilizing 13 m antennas for S-band (2.6–3.95 GHz) communications.

The **GMS** is responsible for the delivery of correct navigation signals and for monitoring system integrity. Therefore, the state of each satellite is continuously monitored in order to ensure that the transmitted signals are within specified tolerances. If deviations outside the specified values are detected, it will be ensured that users will receive information about the defective or missing signals within 6 seconds. The transmitted navigation signal is composed of the accurate time derived from the on-board atomic clocks, the precise satellite orbit data and the integrity information.

The necessary monitoring data is generated using a network of over 40 **Galileo sensor stations (GSS)**, which permanently track all visible Galileo L-band signals with special reference receivers. This data stream is centrally collected and processed in combination with further information (e.g., accurate timing signals from ground station atomic clocks, ionosphere data). A specific GMS computer then calculates new orbit and clock data for each satellite. In addition, the synchronization of all on-board clocks with the GCC clocks, an orbit forecast valid for the next few hours and an integrity assessment for all satellites are performed. This computation is accomplished every 10 minutes. The results of the computations are transmitted to the satellites over nine **uplink stations (ULS)** using 3 m antennas and C-band signals.

Further installations located at the two GCCs control the interfaces to external service providers, monitor the status of the ground elements and provide for the safety and security of the transmitted data.

The individual elements of the Galileo ground segment are described in more detail below.

7.3.4.1 Galileo Ground Control Segment

The part of the Galileo ground segment which is responsible for the operation of the satellite platforms and the maintenance of constellation status is the ground control segment (GCS). The GCS elements are mainly located in the two GCCs or are at the TT&C remote sites. These **TT&C remote sites** ensure the primary contact between the GCCs and the Galileo satellite and are used for uplinking of telecommands (TCs) and receiving telemetry data (TM). The GCS has both real-time and nonreal-time functions.

The GCS elements used for **real-time processing** are:

- Telemetry, tracking and control facilities (TTCFs)
- Satellite data distribution network (SDDN)
- External data distribution network (S-EDDN)
- Spacecraft and constellation control facility (SCCF) including site data handling set (SDHS)
- GCS key management facility (GCS-KMF)
- Central monitoring and control facility (CMCF).

The GCS elements **without real-time requirements** are:

- Satellite constellation planning facility (SCPF)
- Flight dynamics facility (FDF)
- Operation preparation facility (OPF)
- Constellation simulator (CSIM)
- Integrated logistics support tools (ILS tools).

Real-Time Elements of the GCS

The individual Galileo GCS elements are as follows.

Telemetry, Tracking and Control Facilities (TTCFs)

Five TT&C stations are planned for Galileo operations, each having one 13 m S-band antenna which can be moved along two axes in order to track satellites

from horizon to horizon. For optimal distribution three stations are located near the equator and one station each is located at higher northern/southern latitudes.

During normal operations the TT&C stations are unmanned. Intervention by staff is only envisaged for exceptional and emergency cases as well as for maintenance and upgrades. The **central monitoring and control facility** (CMCF) monitors and operates all functions of the TT&C stations. These stations transmit the control signals for the satellites; they are generated at the GCC site by the spacecraft and constellation control facility (SCCF).

The **TC uplink** is used for the transmission of satellite control commands, avionics software patches and maintenance parameters. The TC uplink frequency is 2225 MHz. In addition, the TC uplink is used as an additional communication channel for sending the navigation data produced by the GMS to the satellite once per satellite orbital period. This function serves as backup, making it possible to renew the navigation data at least every 14 hours in case the C-band uplink of the GMS malfunctions.

The **TM downlink** transmits telemetry data, status information and on-board computer dumps from the satellites to the GCS elements. The TM downlink frequency is 2048 MHz.

Both the TC uplink and TM downlink transmit the signals in standard **spread spectrum modulation**. If needed, both channels can also switch to phase-modulated (ESA standard) signals. Communication between the TT&C stations and the GCCs takes place via the satellite data distribution network (SDDN).

Additionally, a ranging determination using **two-way propagation time measurements** of the S-band TT&C signals is planned for generating orbital position data. This data, however, will not be used for the real-time system operations. The ranging results will be stored at each TT&C station and upon request the data will be transmitted to the flight dynamics facility (FDF) to be used for off-line analysis.

Communication Networks (SDDN and S-EDDN)
There are two network categories for the communication of GCS data between the GCCs, to and from the TT&C stations, and with the external GCS interfaces:

- Satellite data distribution network (SDDN)
- External data distribution network (S-EDDN).

Both networks are to be considered as virtual networks. The physical realization of the communications network is called the Galileo data distribution network (GDDN), which transmits both GCS and GMS data.

The SDDN transmits data between the two GCCs and to the TT&C stations. The S-EDDN serves as a communication interface with external installations and services, such as possible LEOP services, or for communications from the satellite manufacturer for software and/or data updates, or for analysis and maintenance purposes. Within the GCCs special local area networks (LANs) are designated to connect the individual GCS elements within the GCCs.

Spacecraft Constellation Control Facility (SCCF)
The SCCF is the core element of the GCS. It processes the telemetry data and generates the appropriate satellite command data for routine operations as well as for special operations. The SCCF can be operated either automatically or manually.

The SCPF generates and conveys the short-term planning for the satellite contact scheduling to the SCCF. The SCCF then produces the real-time commands for satellite control. This SCCF contact and configuration data is then forwarded to the CMCF for the appropriate control of the respective ground elements.

The communication with the TT&C stations is accomplished with a special protocol (CCSDS space link extension), which is generated by an SCCF subsystem, the so-called site data handling set (SDHS). The SDHS is connected to appropriate cryptography modules for ciphering and deciphering the TT&C data.

GCS Key Management Facility (GCS-KMF)
The GCS-KMF is responsible for ciphering and deciphering all TM/TC data. The GCS-KMF is also responsible for controlling the on-board security units.

Central Monitoring and Control Facility (CMCF)
The CMCF monitors and controls all GCS equipment (with the exception of the GCS-KMF, which is

only monitored). Additionally, local monitoring and control units are installed in each TT&C station and are controlled and monitored by the CMCF over the SDDN. The consolidated TT&C data is preprocessed by the CMCF and forwarded to the SCCF for overall coordination and synchronization of the satellite operations. Normally the CMCF is operated automatically; however, it has manual access features for special operations and maintenance.

The CMCF transmits aggregated data to the ground asset control facility (GACF), which is part of the GMS. This is because the overall monitoring and control functions of all Galileo ground elements are performed and coordinated in the GACF.

GCS Elements without Real-Time Requirements

Spacecraft Constellation Planning Facility (SCPF)
The SCPF performs short-term planning for the GCS tasks. SCPF planning includes the operational activities for constellation control and is responsible for scheduling the satellite contact time slots. For this task it is required that the data exchange requests between the 30 satellites and the five TT&C stations be performed with an optimum schedule. To this end the SCPF is subordinate to GMS planning and communicates with the GMS data sets relevant for medium-term planning, for example payload maintenance, satellite maneuvers or software update procedures. This also results in additional requests for TT&C contact time slots.

Flight Dynamics Facility (FDF)
The FDF calculates forecasts for satellite orbits and attitudes and supports the planning for satellite control maneuvers. This serves for both the control of individual satellites as well as the overall management of the entire constellation.

Normally the FDF uses satellite orbit data provided by the GMS. Additionally, the FDF can use for its orbital computations data coming from the two-way S-band propagation time measurements provided by the TT&C stations. This allows satellite control (e.g., for special operations or in emergency situations) completely independent of the availability or functioning of the GMS.

Operation Preparation Facility (OPF)
The OPF with its special editors covers the development and maintenance of operational data and procedures. Additional elements of the OPF guarantee a consistency check of the configuration data and support the import and export of software elements for the satellites' on-board computers.

Constellation Simulator (CSIM)
The CSIM provides a comprehensive simulation tool for the entire constellation and ground systems including the TT&C stations. The CSIM is used to validate operational data and procedures, to investigate anomalies and also for test and training purposes. Each satellite is described by a state model that reacts to TC data input and returns via TM an appropriate data output. The simulator is able to process several scenarios in parallel which reflect different constellation situations.

Integrated Logistics Support Tools (ILS Tools)
The complex Galileo infrastructure must be carefully supervised and maintained during its complete operational lifetime. For this logistics task so-called ILS tools are made available. Their main functions are arranged as follows:

- Logistics management information system (LMIS)
- Central document management system (CDMS)
- Integrated logistics support database
- Logistics support analysis (LSA) development tools.

The ILS tools are not a direct part of the GCS elements and therefore are hosted on separate servers; they are accessible to operations personnel via the administrative LAN of the GCCs.

GCS Interfaces to External Facilities

The GCS interfaces to other facilities and/or service providers are illustrated in Figure 7.3.2.

In particular the data exchange between GCS and GMS is important. This data exchange is performed by the GCC's **internal LAN**.

Data exchange with **external service provider**s, such as an external LEOP service provider or the

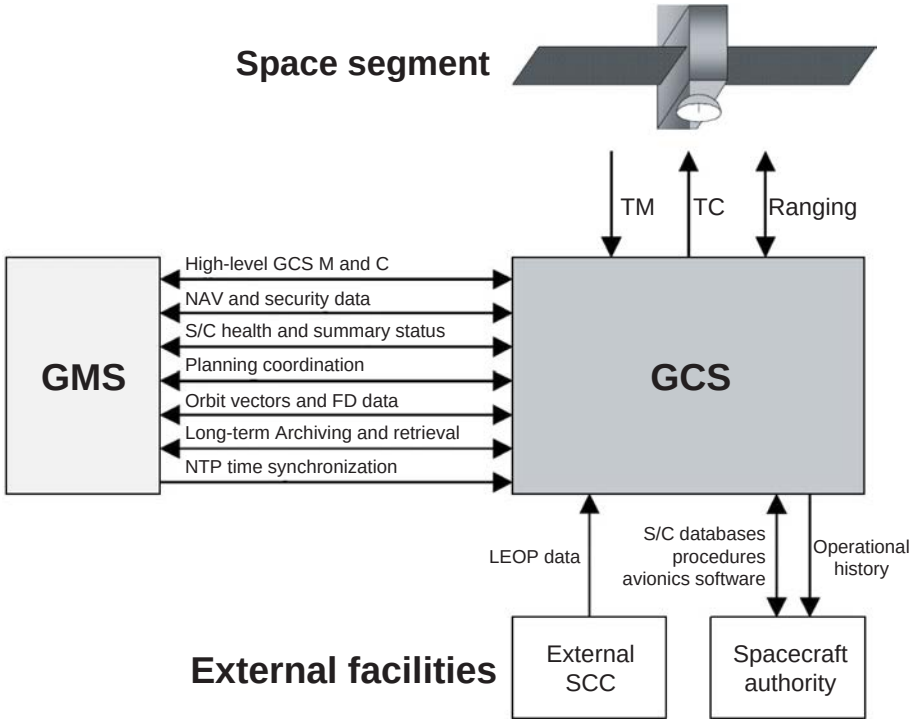


Figure 7.3.2: *Overview of functions and interfaces of the Galileo segments.*

satellite manufacturer for software and/or database updates or for analysis and maintenance, is provided via the E-SDDN.

The **interface to the space segment** is provided by the TT&C stations, which are connected to the GCS by the SDDN. The physical implementation of the communications network which transmits the SDDN data is called the global distribution data network (GDDN), used by GCS and GMS simultaneously.

7.3.4.2 Galileo Ground Mission Segment

GMS Overview and Tasks

The Galileo ground mission segment (GMS) is responsible for the correct generation and distribution of the navigation data. Therefore the GMS is connected to the GCS in order to coordinate all operational activities and to monitor all ground segment installations (see above descriptions for details). Additionally, the GMS is connected to Galileo-external facilities in order to exchange, for example, reference data.

The GMS facilities are installed either in the GCC or at a Galileo sensor station (GSS) or at the uplink station (ULS). The following provides an overview of the GMS facilities, namely:

- Galileo sensor station (GSS)
- Precise time facility (PTF)
- Orbit and synchronization processing facility (OSPF)
- Integrity processing facility (IPF)
- Message generation facility (MGF)
- Service products facility (SPF)
- Ground assets control facility (GACF)
- Mission and uplink control facility (MUCF)
- Mission support facility (MSF)
- Mission key management facility (MKMF)
- Key management facility (PKMF PRS)
- Uplink station (ULS)
- Mission data dissemination network (MDDN).

GMS Internal Facilities

The Galileo GMS elements which are directly assigned to the GMS are briefly described below.

Galileo Sensor Station (GSS)
The GSSs form a global network of (as presently planned) 40 monitoring stations which track all visible satellite navigation signals and transmit the respective data to the GMS facilities in the GCC. A GSS consists of the following elements:

- A receiver, which can receive and demodulate all four frequencies (L1, E5a, E5b and E6).
- An atomic clock and a local frequency generation/dissemination unit, which is used as local reference oscillator.
- A communication interface and an electrical power supply unit.
- A data processor and a data memory.

Precise Time Facility (PTF)
The PTF is sometimes called the "heart of Galileo." It generates the Galileo system time (GST) and basically performs the following tasks:

- Generation of the GST standard by averaging the Galileo clock ensemble data.
- Physical realization of the GST.
- Bringing the GST closer to International Atomic Time (TAI) using the external input of the time service providers (TSPs).
- Computation of the time difference between GST and GPS time.
- Synchronization between the two GST, which are generated in the two GCCs.

Orbit and Synchronization Processing Facility (OSPF)
The OSPF produces the essential core data which is required for the users' navigational computation and which is transmitted to them by the satellites. The OSPF data set comprises:

- Precise ephemeris data for all satellites of the Galileo constellation, given in Galileo terrestrial reference frame (GTRF) coordinates.
- Parameters for modeling the satellite's on-board clocks.
- Parameters for signal error margins, the so-called "signal in space accuracy" (SISA).
- Parameters for the satellite almanac.
- Parameters for single-frequency users (e.g., used for ionosphere corrections).

Integrity Processing Facility (IPF)
The IPF produces the integrity information for users of the SoL and PRS services, which indicate in near real time whether the signal is within the specified limits. The IPF produces the integrity flag (IF) and the observable SISMA (Signal In Space Monitoring Accuracy), which are transmitted to the MSF for dissemination to the users.

Message Generation Facility (MGF)
The MGF collects all data to be transmitted to the satellites over the C-band uplink. From the complete data set the MGF produces the appropriate data packages, which then are dispatched to the ULS. Likewise, the MGF produces the data packages which are dispatched over the GCS S-band uplink.

Service Products Facility (SPF)
The SPF produces special information which is made accessible outside of the Galileo infrastructure, namely:

- Information about the Galileo operational status and usability, which can be accessed via the Internet.
- Information which is accessible over the Internet via a programmed user-specific interface.
- Automatically generated event-triggered information, such as regular reports, special or status reports, maintenance information.

Ground Assets Control Facility (GACF)
The GACF is the central core which monitors, coordinates and controls all Galileo ground infrastructure (GMS and GCS) facilities. GMS maintenance is also controlled via the GACF. Additionally, the GACF hosts the central GMS archive.

Mission and Uplink Control Facility (MUCF)
For management of the GMS operations, the MUCF establishes an overview of all GMS processes. It has the following components/functions:

- Galileo mission and scheduling for the GMS functions regarding medium-term and long-term planning.
- Establishing, storing, validating and transmitting (to the MGF and ULS) the ULS contact time schedule.

- Supervision of all GMS tasks.
- A posteriori verification and an a priori forecast of navigation performance data.

Mission Support Facility (MSF)
The MSF is an off-line facility for the GMS engineers and is used for the following tasks:

- Computation of reference data for satellite orbits and clock error determination, in order to support the fine tuning of the OD&TS algorithms and to determine calibration parameters.
- Reference computation of the integrity and SISA data.
- Monitoring of the signal spectra which are transmitted by the satellites.
- Trend analyses for prediction and maintenance purposes.
- Support for GMS off-line training and upkeep purposes.
- Support functions during GMS deployment and test phases.

Mission Key Management Facility (MKMF)
The MKMF belongs to the Galileo safety and security architecture, which is responsible for the key management and protection of the data for SoL authentication, for protecting the commercial service (CS) data and for C-band communication.

PRS (Public Regulated Service) Key Management Facility (PKMF)
The PKMF is another facility of the Galileo safety architecture which is responsible for the key management and the protection of the PRS data. Both the MKMF and the PKMF are installed in the GCC building in a specially secured zone in order to avoid unauthorized access. For safety and security reasons the details of the MKMF and the PKMF are also subject to classification.

Uplink Station (ULS)
At present, nine ULS sites each having four C-band antennas with a diameter of 3 m are planned for the Galileo ULSs. The ULSs perform the following functions:

- Receipt of data produced by the MGF.
- Pointing of the selected antennas to the predetermined satellite according to the schedule provided by the MUCF.
- Generation of the carrier frequency for C-band communication, modulation of data to the baseband, and transmission of the signals to the satellites.

The uplink frequency in the C-band will be 5005 MHz.

Mission Data Dissemination Network (MDDN)
The MDDN is responsible for communicating the GMS data between the GCC and the GMS remote sites (i.e., GSS and ULS sites). The data flow is as follows:

- Transfer of the navigation raw data generated by the global GSS network to the GCC.
- Transfer of the GMS data to the ULSs.
- Transfer of data between the GCCs.
- Communication with the external facilities and service providers interfacing with GMS elements.

The GDDN is the physical realization of the data communication interconnections.

External GMS Facilities
During Galileo operations the GMS elements interface with the following external GMS elements, facilities and service providers:

- Galileo space segment (30 MEO satellites)
- Galileo ground control segment (GCS)
- Geodetic reference service provider (GRSP)
- Time reference service providers (TSPs)
- Service center
- Regional integrity data supplied by external entities (ERIS)
- GPS navigation system.

The first two interfaces are described above; the other interfaces are described in the following.

Geodetic Reference Service Provider (GRSP)
All Galileo position data is defined in Galileo's own geodetic reference frame, the so-called Galileo

terrestrial reference frame (GTRF). External service providers, the GRSPs, regularly supply data which refers to geodetic references. These are, for example, the position changes of the GSS stations measured in the ITRF (International Terrestrial Reference Frame), or changes in the Earth's rotation axis or rate.

The consistency between GTRF and ITRF is also guaranteed by the GRSP and adjusted on a long-term basis with all other data of the International Earth Rotation Service (IERS).

Additionally, the GRSP is to supply data from laser ranging measurements between the ground station and satellites; these allow in individual cases orbit determination for measurement and calibration purposes to millimeter accuracy.

Time Reference Service Providers (TSPs)
The Galileo GMS generates the so-called **Galileo reference system time** (GST). As per Galileo mission requirements, GST should be synchronized with TAI. The maximum offset between GST and TAI is specified to be 50 ns for 95% of any time period in one year. Thus the offset is to be determined to an accuracy of less than 28 ns with a confidence of 95%.

The uncertainty in the **offset** between GST and TAI results mainly from the uncertainty of the TAI prediction, since TAI can be determined only retrospectively and publication takes place only once per month. In order to fulfill the GST requirements with respect to TAI synchronicity, the so-called TSPs (usually UTC time laboratories) supply steering parameters to the GMS. The TSP is assigned the following tasks:

- Determination of a daily estimated value of TAI (predicted) – GST and establishment of the respective frequency offset.
- Daily generation of a set of parameters to steer the system as per the GST requirements.

Service Center
Beyond the signal in space (SIS), Galileo users will have additional ways to interface with the Galileo system. The service center provides specific, off-line data services, access to archived data, time-dependent variations to model parameters, etc. The service center has not yet been fully defined.

Regional Integrity Data Provided by External Suppliers (ERIS)
ERIS provides the possibility for regionally generated integrity data to be directly or indirectly transmitted to the Galileo satellites in order to broadcast it over the respective regions.

In the case of direct transmission of the ERIS information, it will be uplinked by the so-called **ERIS uplink station** (EULS) to one of the six C-band channels on-board the satellites.

In the case of indirect transmission, the ERIS information will be sent to the GMS and then forwarded to the respective satellites.

GPS Navigation System
In order to guarantee **interoperability** between Galileo and GPS, that is to provide the possibility for users that the receivers can simultaneously calculate a position with a mix of Galileo and GPS signals, the time offset between Galileo and GPS system time (GGTO) is transmitted as part of the Galileo navigation message. Therefore, a standardized two-way time synchronization called two-way satellite time and frequency transfer (TWSTFT) between the Galileo GMS and the GPS ground segment will be implemented. Additionally (as backup), combined GPS/Galileo receivers will be used as data sources for GGTO determination.

7.3.5 Navigation Signals and Services

7.3.5.1 Reference Systems for Time and Coordinate Frames

Time
Today the standard (SI) unit of time is the **second**, which is defined by the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom [7.3.8].

TAI (Temps Atomique International) is used as the international time scale and the second is based on TAI. The reference point in time is January 1, 1970. The physical realization of atomic time is generated by the weighted mean value of about 300 atomic clocks located at about 65 different time laboratories, such as the National Physical Laboratory (NPL) in the UK. The

SI second is significantly more accurate than the second which is derived from a mean sidereal day, because of both the fluctuation of the Earth's rotation rate and its gradual deceleration. Therefore an offset results between the SI second and sidereal time. Normally atomic time and TAI are not directly used. Instead, **Coordinated Universal Time** (UTC) is of practical relevance. As with the atomic time standard, UTC is based on the SI second. The relation between UTC and the various national real-time realizations (e.g., UTC (NPL)) is published posteriorly by the Bureau International des Poids et Mesures (BIMP) in monthly bulletins. The deviation between UTC and universal sidereal time (derived from the true phase angle of the Earth's rotation, i.e., the angle between the Greenwich meridian and vernal equinox) is limited to 0.9 seconds by the introduction of a leap second. Since the last introduction of a leap second at the turn of the year 2008/2009, the difference amounts to TAI − UTC = 34 s.

GPS time (GPST) is determined from a clock ensemble composed of the cesium and rubidium atomic clocks at the GPS ground stations and the atomic clocks on-board the GPS satellites. As with UTC, GPST is also a weighted mean average time, but with two substantial differences. GPST is available in real time and is a continuous time without leap seconds. It is steered to be as synchronous as possible with UTC. In recent years the deviation from UTC (UNSO) was within about 10 ns (modulo 1 second). Since January 2009 the difference between UTC and GPST amounts to GPST − UTC ≅ 15 s.

This time difference is transmitted in the GPS navigation data message, so that each receiver can calculate UTC.

The first epoch of GPST started at midnight between Saturday, January 5, and Sunday, January 6, 1980; hence 00:00:00 UTC 06 January 1980. The GPST epoch is defined as a week number and a second within the current week, starting with the night of Saturday to Sunday.

Since the GPS week is represented by 10 bits in the navigation message, the first overflow and reset to week 0 took place between August 21 and 22, 1999. Every 6 seconds the number of seconds in the current epoch is transmitted within each GPS satellite's navigation data message as a time stamp in synchrony with the satellite clock. The GPS master control station (MCS) controls all clocks on-board the GPS satellites and selects one per satellite for signal generation. The deviation between the selected clock and GPST is modeled with a quadratic function of time, and the parameters of this model are calculated in the MCS and transmitted as part of the broadcast navigation data message. The **deviation** $\delta t^s = t^s - t$ at the point in time t is calculated as follows:

$$\delta t^s(t) = a_0 + a_1 (t - t_0) + a_2 (t - t_0)^2 + \Delta t_r \qquad (7.3.6)$$

Here t_0 marks the reference epoch, a_0 is the clock offset (in seconds), a_1 is the relative frequency instability (in seconds/second) and a_2 is the relative frequency shift (in seconds/second²). The parameters are often called clock bias, drift and aging. The parameter a_0 is typically between 1 μs and 1 ms. Parameter a_1 is about 10^{-11} s/s. In the last few years it was frequently set to zero; $a_2 = 0$ s/s².

The term Δt_r considers the necessary relativistic time adjustment. With these adjustment parameters, which are usually calculated once per day and then transmitted to the satellites, the satellite clock errors are within about 5–10 ns.

According to **Einstein's theory of relativity**, a clock's frequency changes depending on its motion and on the change in gravitational potential. The clocks on-board satellites are about 20 000 km above the geoids (which to a good approximation is identical to mean sea level) and therefore show a higher frequency, as according to the general theory of relativity. Due to the satellite's velocity in orbit of about 4 km/s, the clocks are slower than clocks on the Earth's surface. At an altitude of about 3000 km these two effects would compensate each other. A clock in a circular orbit with a radius of 26 560 km would gain 38.4 μs per day compared to a clock on the Earth's surface. In order to compensate for this, the GPS on-board clocks are not tuned to the nominal frequency of 10.23 MHz, but are set to a different frequency, which is reduced by the so-called "factory offset" of 0.004 567 3 Hz.

Since the GPS orbits are not perfectly circular, but elliptical, the deviations resulting from relativistic effects must be corrected by the receivers. The accurate orbit parameters are transmitted once per day from the MCS to the satellites and are continuously broadcast as part of

the navigation message. The interface control document IS-GPS-200-Rev-D [7.3.10] describes the algorithm for the time adjustment in the receiver, which can amount to 45 ns, depending on the orbital position.

Coordinates Frame: WGS84, ITRF and GTRF

The World Geodetic System 1984 (WGS84) is the geodetic basis for GPS and for global measurements with NAVSTAR GPS satellites [7.3.9]. As part of the WGS84 standard, a global **reference spheroid** which best approximates the Earth's surface is defined. The term WGS84 represents the datum, the fundamental system and the reference spheroid, which is called GRS80. The WGS84 standard has been taken over by EUROCONTROL in the meantime also for aviation.

In addition to this global reference frame, numerous local reference frames are used. In Germany, for example, a local reference ellipsoid (Bessel ellipsoid, named after Friedrich Wilhelm Bessel) with its fundamental point in Rauenberg is known as the Potsdam Datum. In the future, the **European Terrestrial Reference System 1989** (ETRS89) will be used as the European standard for land survey.

ETRS89 is a three-dimensional geodetic reference system. It was defined by the European Subcommittee of the IAG (EUREF). It was fixed to the stable part of the Eurasian continental plate and is identical to ITRS at epoch 1989.0. ETRS89 is realized by EUREF through a sporadically published ETRF_YY (European Terrestrial Reference Frame, with designation year) and specifies **three-dimensional coordinates and velocities** for selected fixed points determined by **global measurement procedures** such as very long baseline interferometry (VLBI), satellite laser ranging (SLR) and by utilizing GPS and Doppler orbitography and radiopositioning integrated by satellite (DORIS). These reference points are used in the definition of the hierarchical level A of the ETRF. During the 1980s and 1990s mobile measurement campaigns took place at ground-marked fixed points. Today, on the European level ETRS89 is generated from permanent GPS measurements at reference sites of the EPN (EUREF Permanent GPS Network). As a result of the long observation times and the stability of the reference sites, which are fixed on solid buildings, an accuracy of a few millimeters is achieved. The velocity of the reference sites represents the Earth's surface movement on the Eurasian Plate. ETRS89 was recommended by the European Union as the uniform reference system for European geodata. At the epoch 01 January 1989 the coordinates from ETRS89 and WGS84 differed by less than 1 m.

7.3.5.2 Navigation Signals

The navigation signals consist of three components: **carrier frequency**, **pseudo-random noise (PRN) code** and **navigation data** [7.3.3], [7.3.4]. Figure 7.3.3 shows this structure and the resulting signals on the basis of the GPS C/A code in the L1-band.

All signals are derived from the primary pulse of the atomic clock on-board the satellites and all signals are synchronized with the on-board clock. The **carrier frequencies** of all satellite navigation signals are in the L-band between 1.2 and 1.6 GHz [7.3.10], [7.3.11]. The so-called **PRN code** is modulated onto the carrier frequency. The PRN code varies for each satellite and is characterized by special mathematical features which permit all signals to be sent in the same frequency band without mutual interference. The autocorrelation function of the signal is maximum, while the cross-correlation function between different satellite signals is minimum. Thus, by using a PRN code replica stored in the receivers, the satellites can be uniquely identified. The PRN code also serves as the basis for propagation time measurements.

Additionally, the so-called **navigation data** is modulated onto the PRN code. The navigation data contains precise orbit data, clock parameters, rough orbit data of all satellites (almanac), ionosphere data and satellite status data. The transmission of the complete navigation message takes 12.5 minutes. On the basis of this information the receiver can determine an accurate satellite position at the time of signal transmission and the respective satellite time (measured in the global system time). The transmission position and time allow calculation of the propagation time of the signal.

In order to fulfill various technical requirements and to make position determination as precise and reliable as possible, the navigation signals are transmitted at **multiple frequencies**. Therefore, a navigation service, such as C/A or P(Y), usually uses

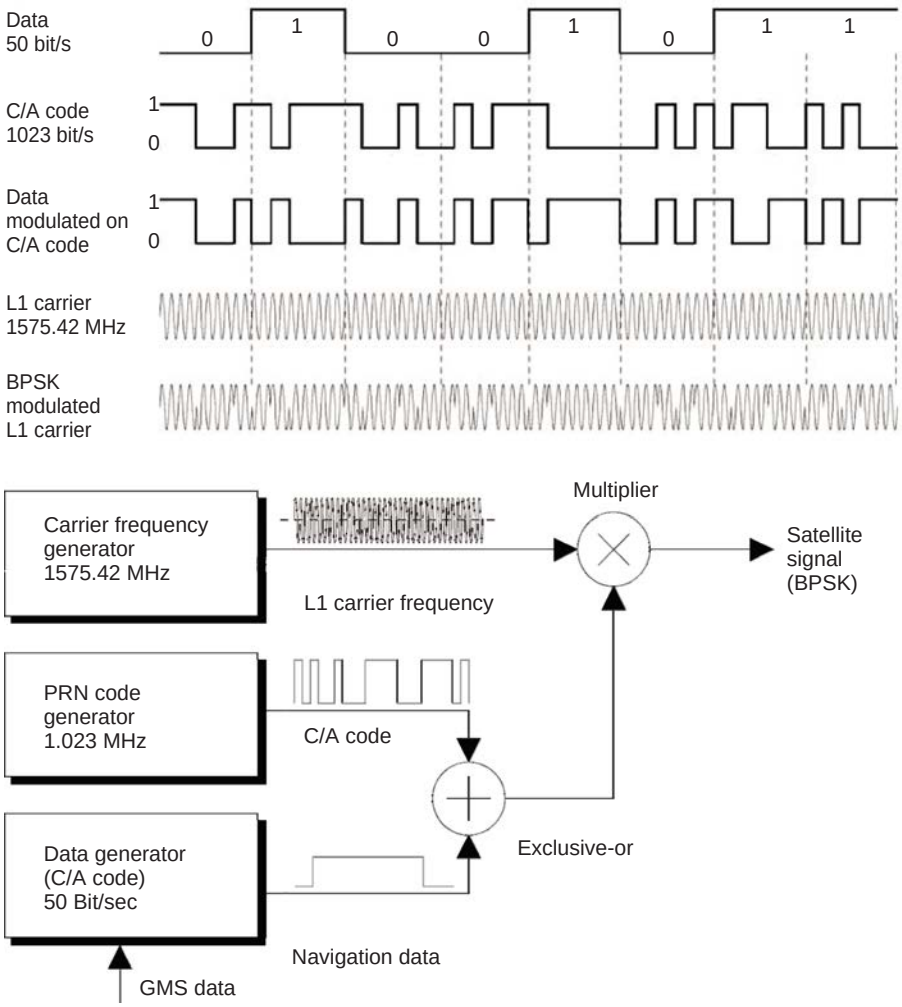


Figure 7.3.3: *Structure, composition and pattern for generating GPS navigation data.*

several frequencies. Table 7.3.4 gives an overview of the different frequencies, the associated modulations and the service attributes.

7.3.5.3 Galileo Service Concept

GPS has two navigation services, a freely available service (C/A) and a governmental encrypted service (P(Y)). For Galileo the concept was expanded to include five services [7.3.6].

The **open service** (OS) results from a combination of open signals, free of user charge, and provides position and timing performance comparable with other GNSS systems.

The **commercial service** (CS) provides access to two additional signals to allow for a higher data throughput rate and to enable users to improve accuracy. The signals are encrypted for commercial access control. It is envisaged that a service guarantee will be provided for this service. It is meant for professional applications, such as geodesy, Earth exploration and network synchronization.

The **safety of life service** (SoL) is planned for groups of users who require guaranteed accuracy as an essential feature. The application areas are to be found in the transport sector (e.g., aeronautical, maritime and rail transport). The Galileo signals will be permanently monitored by a global network of sensor stations. The

Table 7.3.4: *Overview of the signals of the GPS, Galileo and GLONASS satellite navigation systems.*

System	Frequency	Band	Modulation	Data	Signal/Service
GPS	1 176,45 MHz	L5-I	BPSK (10)	1000 sps	C/A-code with Block II-F
		L5-Q	BPSK (10)	Pilot	C/A-code Pilot
	1 227,6 MHz	L2-C	BPSK (1)	50 sps	C/A-code moderate
		L2-P	BPSK (10)	Pilot	C/A-code long
		L2-M	BOC (10,5)	Encrypted	M-code
	1 575,42 MHz	L1-C/A	BPSK (1)	50 sps	C/A-code
		L1-P	BPSK (10)	50sps Encrypted	P(Y)-code
		L1-M	BOC (10,5)	Encrypted	M-code
Galileo	1 176,45 MHz	E5a-I	AltBOC (15,10)	50 sps	OS/CS/SoL
		E5a-Q	AltBOC (15,10)	Pilot	OS/CS/SoL Pilot
		E5b-I	AltBOC (15,10)	250 sps	OS/CS/SoL
		E5b-Q	AltBOC (15,10)	Pilot	OS/CS/SoL Pilot
	1 278,75 MHz	E6A	BOCcos(10,5)	Encrypted	PRS
		E6B-I	BPSK (5)	1000 sps	CS
		E6B-Q	BPSK (5)	Pilot	CS Pilot
	1 575,42 MHz	L1A	BOCcos(15,2.5)	Encrypted	PRS
		L1B	BOC (1,1)	250 sps	OS/CS/SoL
		L1C	BOC (1,1)	Pilot	OS/CS/SoL Pilot
GLONASS	1 201,5 MHz	G3	BPSK (4)	Encrypted	P-code
		G3	BPSK (4)	Encrypted	P-code
	1 246,0 MHz	G2	BPSK (5.11)	Encrypted	P-code
		G2	BPSK (0.511)	50 sps	C/A-code
	1 602,0 MHz	G1	BPSK (5.11)	Encrypted	P-code
		G1	BPSK (0.511)	50 sps	C/A-code

7

monitoring data will be transmitted to the Galileo control center and a so-called integrity signal will be generated and broadcasted by all Galileo satellites as part of the SoL data message. Any event or navigation signal outside the specifications will be transmitted to users by the integrity signal within 6 seconds.

The **public regulated service** (PRS) is encrypted, has governmental-controlled access and is more resistant to unintentional or deliberate interference. The service is provided for governmental users such as police, customs and national security organizations. PRS serves sovereign tasks of the member states of the European Union. PRS must be permanently operated and will be available under all circumstances, which includes in particular hazardous situations or times of crisis.

The **search and rescue** (SAR) service is not a pure navigation service, but allows the reception of emergency calls from any position on Earth practically in real time. The service supports already existing SAR systems like COSPAS-SARSAT. The SAR uplink frequency (from the user to the satellite) is at 406 MHz. The SAR downlink frequency (from the satellite to the SAR control center) is at 1544 MHz. The SAR service also provides for a feedback channel to persons in distress.

7.3.6 Receiver

Since the beginning in about 1980 and driven by vigorous developments in microelectronics, the

capability and efficiency of integrated GNSS receivers have substantially increased while the size and power consumption have noticeably decreased. However, the basic operational principle of GNSS receivers remains essentially unchanged. This principle is described using the example of a simplified GPS receiver structure.

The **antenna** receives L-band (1.2–1.6 GHz) signals. A radiofrequency preamplifier followed by a downconverter stage makes the signals available in baseband.

Each GPS satellite is characterized by a specific PRN code. The different signals from different satellites can therefore be separated and tracked using a replication of the PRN code, which is stored in the receiver. To this end the input signal is multiplied by the stored code replication. For this process the code replication is shifted along the time axis until the maximum of the correlation function is reached. This time difference is transformed into a ranging value, which is called the **pseudo-range**. The word "pseudo" refers to the fact that it does not reflect the real distance between receivers and satellite, but only a receiver-internal variable.

The data of the **navigation message** (ephemeris and parameter of the satellite clock) is used to compute position, velocity and time, in combination with at least three additional **pseudo-range measurements**. With the navigation message data and at least four pseudo-ranges it is possible to derive three position coordinates and the time. Figure 7.3.4 shows the respective signal flow within a receiver.

This simple receiver structure is adapted and/or extended depending upon its application and requirements. Thus dual-frequency receivers have been developed in order to correct ionospheric propagation time variations. Recently, triple-frequency receivers have also been developed and are available.

For applications requiring the **highest precision** a class of receivers has been developed which in addition to the code phase measurements also utilizes L-band carrier phase measurements. Due to the short wavelength of about 20 cm it is possible to achieve a relative accuracy of better than 5 cm, which is known as the carrier phase measurement or RTK measurement. For this type of measurement the error sources (such as orbit error, clock error, etc.) must be cancelled by differential correction technique and the phase ambiguity must be resolved.

For users requiring a maximum number of visible satellites (e.g., RTK measurements under reduced satellite visibility in urban canyons), multiconstellation receiver types have been developed with simultaneous GPS and GLONASS signal processing capability. Combined **GPS/Galileo receivers** are available in prototype versions.

7.3.7 Accuracy and Error Factors

The positioning is subject to numerous noise influences. Figure 7.3.5 shows an overview of the substantial interference sources and the typical orders of magnitude of these faults.

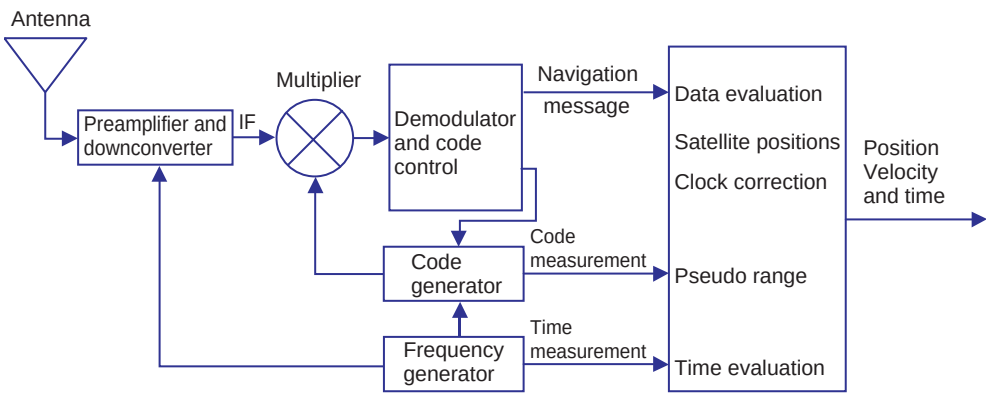


Figure 7.3.4: *Signal flow diagram for a GPS receiver.*

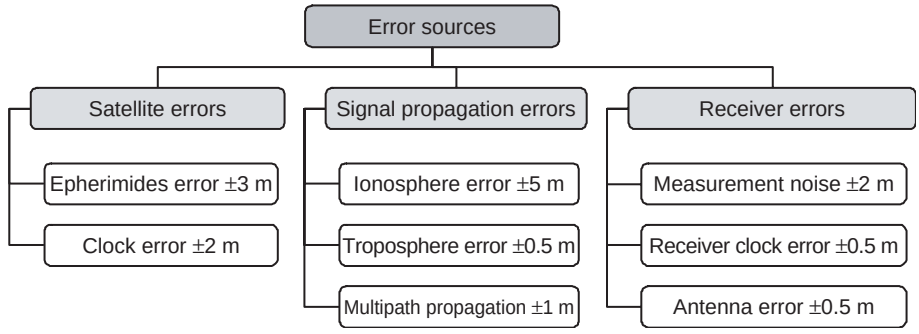


Figure 7.3.5: *Error sources of navigation signals.*

7.3.7.1 Satellite Sources of Error

Since both the satellite orbits and the satellite time are described by parameters in the navigation message and these parameters can "age" up to 12 hours, the actual track and clock data deviates from the estimations.

This error can be substantially reduced by so-called **differential correction**. At accurately aligned reference points, the deviations being produced are determined and transmitted to the user, for example by means of portable radio or satellite communication. As a function of the latency of the adjustment data and the distance from the reference station, the satellite error can be reduced to below a few centimeters.

7.3.7.2 Signal Propagation Error

The propagation velocity of the signals from the satellite to the ground receiver can be altered by the terrestrial atmosphere. A distinction is made between ionosphere- and troposphere-introduced errors.

The concentration and distribution of free electrons in the **ionosphere** can vary considerably, for instance due to changes in the solar wind. These fluctuations can sometimes occur on small scales where the diameter of local ionospheric turbulence is under 50 km. The runtime fluctuations can cause errors of up to several meters. By using two-frequency receivers, the propagation time effects can be modeled in various frequency bands and a significant part of this error can be corrected. Nevertheless, at times of high solar activity substantial limitations on the measuring accuracy can be observed.

The humidity of the **troposphere** (clouds, rain) likewise leads to altered signal propagation times. Usually the error caused here is less than about 1 m.

Signals which did not arrive in the direct line of sight of the receiver but were reflected by the environment of the receiver (e.g., by houses, objects in the vicinity or on the ground) show a deceptive position. This effect is called **multipath propagation**. Numerous receiver technologies have been developed to suppress this positioning error. For example, the signal shape (changes are caused by rapid phase jumps during reflection), or an evaluation of only the directly received signals and a suppression of delayed signals, can be used to identify and reduce the multipath effect in the receiver.

7.3.7.3 Receiver Errors

Since the satellite signals are extremely weak (10^{-16} W), the **thermal noise** of the receiver substantially limits the proper evaluation of the signal [7.3.3]. Temporal correlation between received signal and stored code replica is possible with a resolution of about 1% of the chip rate of the signal, which corresponds to a GPS C/A code resolution of about 10 ns and/or about 3 m. With the GPS P(Y) code the chip rate is 10 times higher, which reduces this error to about 0.3 m, making it negligible when combined with the other sources of error.

Depending upon receiver cost and design, the phase noise of the receiver clock and changes in the antenna's center may be significant factors in error analysis.

Bibliography

[7.3.1] Williams, J.E.D. *From Sails to Satellites: The Origin and Development of Navigational Science*. Oxford: Oxford University Press, 1994.

[7.3.2] Born, D.J. *Crosslink – Satellite Navigation*, Vol. 3, No. 2. Los Angeles: The Aerospace Corporation, 2002.

[7.3.3] Misra, P., Enge, P. *Global Positioning System, Signals, Measurements, and Performance*, Second Edition. Lincoln, MA: Ganga-Jamuna Press, 2006.

[7.3.4] Kaplan, E.D., Hegarty, C.J. *Understanding GPS: Principles and Applications*, Second Edition. London: Artech House, 2005.

[7.3.5] Battrick, B., Danesy, D. *EGNOS – The European Geostationary Navigation Overlay System*. Noordwijk: ESA Publications Division, 2006.

[7.3.6] Hein, G.W. *et al.* Envisioning a Future GNSS System of Systems. Parts 1–3. *Inside GNSS*, Vol. 2, Nos. 1–3, 2007.

[7.3.7] Montenbruck, O., Gil, E. *Satellite Orbits: Models, Methods and Applications*. Heidelberg: Springer Verlag, 2001.

[7.3.8] Allan, D.W., Ashby, N., Hodge, C.C. The Science of Timekeeping. *Hewlett Packard Application Note 1289*, 1997.

[7.3.9] NIMA TR8350.2. *Department of Defense World Geodetic System 1984: Its Definition and Relationships with Local Geodetic Systems*, Third Edition. National Imagery and Mapping Agency, 1997.

[7.3.10] IS-GPS-200D. *Interface Specification – Global Positioning System*. El Segundo, CA: NAVSTAR GPS Joint Program Office, 2004.

[7.3.11] IS-GPS-705D. *Interface Specification – Global Positioning System*. El Segundo, CA: NAVSTAR GPS Joint Program Office, 2005.

7.4 Space Astronomy and Planetary Missions

Stephan Ulamec

Since early history, humans have wondered about the **nature of the objects that can be observed in the nocturnal sky**. Space activities have opened up entirely new possibilities in astronomy and also for the exploration of outer space.

Ground-based astronomical observations are restricted to those **wavelength ranges** for which the terrestrial atmosphere is transparent. This restriction, obviously, no longer applies to satellites, which allow observations without the disturbing influence of the atmosphere. This refers to X-ray, gamma, ultraviolet and infrared astronomy. But the atmosphere also affects observations in the visible range, which is why, for example, the Hubble Space Telescope (HST) allows much better observations compared to the (sometimes much larger) ground-based telescopes.

The second aspect of space activity is the **exploration of our Solar System**. Since the 1960s the planets of the Solar System have been "visited" by space probes, first in the course of fly-by missions, later with orbiters and landers. The study of Earth's Moon has played a particularly important role.

7.4.1 Astronomy Missions

Figure 7.4.1 shows the **opacity** (1/transmission) of the Earth's atmosphere as a function of wavelength. It indicates that observations in the visible range in the near ultraviolet and infrared, as well as in the radio wave range, are possible from the ground. However, the electromagnetic spectra of shorter wavelengths as well as of lower energy infrared cannot be observed from the ground and could only be "discovered" with satellites.

7.4.1.1 X-ray Satellites

The first satellite to discover **cosmic X-rays** was **OSO-3** (Orbiting Solar Observatory) in 1960. OSO-3 was actually designed to observe the Sun; however, it also detected diffuse cosmic X-rays.

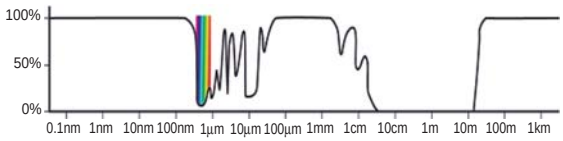


Figure 7.4.1: *Opacity of the Earth's atmosphere as a function of wavelength (Source: NASA).*

The first mission dedicated to the search for cosmic X-ray sources was **Uhuru**, launched in 1970. Uhuru provided the first X-ray map of the sky in the energy range between 2 and 20 keV.

Many missions followed, two examples are briefly described below: ROSAT and XMM-Newton.

ROSAT (Roentgen Satellite), a German–US–British joint effort, was launched in June 1990 and was in operation for nine years until February 1999. Its sensors were about 1000 times more sensitive than, for example, those of Uhuru. The payload consisted of a telescope, which could be combined with either a position-sensitive proportional counter (PSPC) or an imaging instrument (high-resolution imager, HRI), as well as a wide-angle camera.

A total of about 150 000 X-ray sources were mapped. Supernova remnants, galaxy clusters, neutron stars and many other phenomena were investigated in detail.

XMM-Newton (X-ray Multimirror), a mission of the European Space Agency (ESA), was launched in 1999 with a nominal service life of 10 years (Figure 7.4.2). For the first time optical and X-ray observations could be combined. XMM is distinguished by its high sensitivity. The energy range covered by the three X-ray telescopes extends from 0.1 to 15 keV.



Figure 7.4.2: *XMM in the ESTEC integration room (Source: ESA/D. Parker).*

7.4.1.2 Gamma Ray Astronomy

Space-bound observations allow investigation of the even higher energetic gamma radiation. The first satellite to do so was **Explorer 11**, launched in 1961. At that time scientists were able to identify 22(!) gamma particles. Subsequent satellites in the 1960s and 1970s (originally intended to detect tests of nuclear weapons) provided among other observations the first hints of the phenomenon of gamma bursts.

In 1991 NASA launched the **Compton Gamma Ray Observatory**, which was in operation until 2000. With its 17 ton payload (until that time the heaviest astronomy payload ever launched) and four different instruments, Compton could cover the energy range up to 30 GeV.

Since 2002 the European gamma ray observatory **Integral** has been in orbit. Integral studies the most energetic and most exotic phenomena in the Universe such as supernovae, neutron stars, black holes, gamma bursts and active galaxy cores. It combines gamma, X-ray and optical observations.

7.4.1.3 Infrared Satellites

Since the 1970s there have been discussions about also investigating the infrared sky with orbiting observatories. The particular challenge of such missions is the fact that the sensors need to be cooled.

In 1983, **IRAS** (Infrared Astronomical Satellite), a joint project of the USA, the Netherlands and the UK, was launched. The telescope itself was integrated into a Dewar container and cooled with liquid helium. The mission duration was about 10 months (until the available helium was used up); 96% of the sky was imaged several times, and over half a million infrared sources were detected. The core of the Milky Way could be imaged for the first time.

In 1995 ESA launched the **European Infrared Science Observatory** (ISO), which monitored the infrared sky in another wavelength range (2.5–240 μm) for two and a half years with a sensitivity about 1000



Figure 7.4.3: *ISO – Infrared Science Observatory (Source: ESA).*

times greater than that of IRAS. ISO discovered, for example, water ice in interstellar dust clouds and hydrocarbons in planetary nebula. Figure 7.4.3 shows an artist's impression of ISO in orbit.

The next step in infrared astronomy will be the **Herschel–Planck** mission (formerly called FIRST) planned to be launched by ESA in 2009. Herschel will cover the wavelength range from far infrared to submillimeter (60 to 670 μm). Planck (to be launched simultaneously with Herschel on an Ariane 5 rocket) will investigate the cosmic background radiation (about 2.7 K).

The first satellite to investigate this extreme long-wave infrared range was **COBE** (Cosmic Background Explorer), launched by NASA in 1989. Theories about the big bang were confirmed; the variations of the cosmic background are extremely small (2.725 ± 0.002 K).

7.4.1.4 The Hubble Space Telescope (HST)

The Hubble Space Telescope (named after the American astronomer Edwin Hubble) should be mentioned in particular because of its enormous influence on our understanding of the Universe.

While observations from ground-based telescopes are always disturbed by the Earth's atmosphere ("seeing"), this problem disappears in outer space. The HST could, thus, provide images of unprecedented sharpness.

Only a few of the many discoveries and observations that were accomplished with the HST are listed below:

- The **age of the Universe** was determined at 13 to 14 billion years.
- **Protoplanetary disks** (solar systems in an early phase of their formation) were imaged.
- **Supernova explosions**, showing unanticipated structures, were detected.
- **Very distant and old galaxies** were observed which allowed a glance at the early epochs of the Universe.
- Valuable data was obtained on the **physics of black holes**.
- The acceleration of the **expansion of the Universe** was measured, indicating the existence of dark energy.

As early as 1923, Hermann Oberth, the German rocket pioneer, was the first to have the idea of launching a telescope into space. NASA's "Large Space Telescope" project has existed since 1969, and the launch of the HST was eventually planned for 1986. However, because of the Space Shuttle accident in that year the launch had to be postponed until April 1990.

Five instruments were on-board:

- Wide-field/planetary camera
- High-resolution spectrograph
- Faint-object camera (an ESA contribution)
- Faint-object spectrograph
- High-speed photometer.

Due to problems with the **optics** at the beginning of the mission, the images provided by the HST were not as sharp as expected.

During an HST servicing mission in 1993 the problem was eliminated by adding a corrective lens system (COSTAR). Other servicing missions took place in 1997, 1999 and 2002. The last of such missions has been approved for 2009.

Figure 7.4.4 shows astronauts Jim Newman and Mike Massimino removing the faint-object camera



Figure 7.4.4: *Astronauts Newman and Massimino removing the FOC from the Hubble Space Telescope on March 7, 2002 (Source: NASA).*

(FOC) to make room for the advanced camera for surveys (ACS). This activity took place in March 2002.

The HST (Figure 7.4.5) is a Cassegrain reflector. The primary mirror has a diameter of 2.4 m and a focal length of 57.6 m. The entire telescope has a mass of 11.6 t and is over 13 m long.

The HST is in a circular orbit 590 km above the Earth's surface. Its inclination is 28.5° (the geographic latitude of Kennedy Space Center).

7.4.2 Moon Missions

Robotic missions to the Moon were the logical first step in exploring our Solar System beyond Earth.



Figure 7.4.5: *The Hubble Space Telescope photographed from the Space Shuttle Discovery (SM3A) (Source: ESA/NASA).*

No attempt will be made here to elaborate on the ancient human dream of "flying to the Moon" or the various conceptions in literature. The first real attempts to send a space vehicle to the Moon were

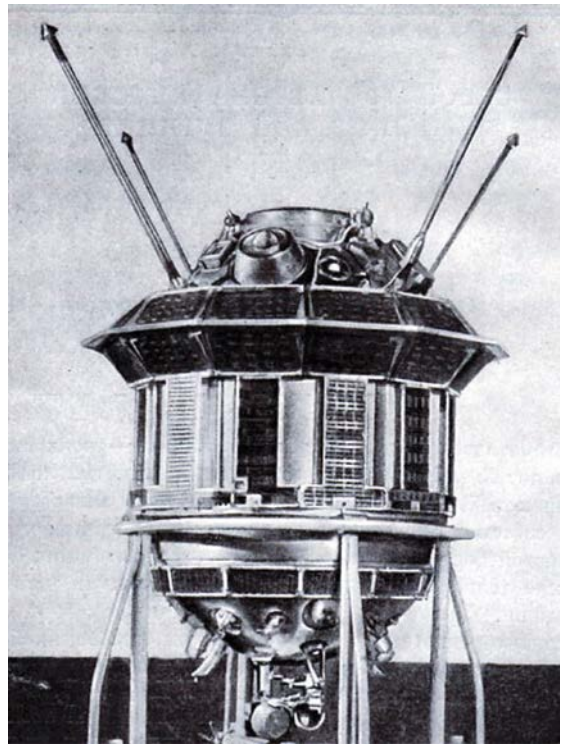


Figure 7.4.6: *Lunik 3 (Source: Lavochkin/NASA).*

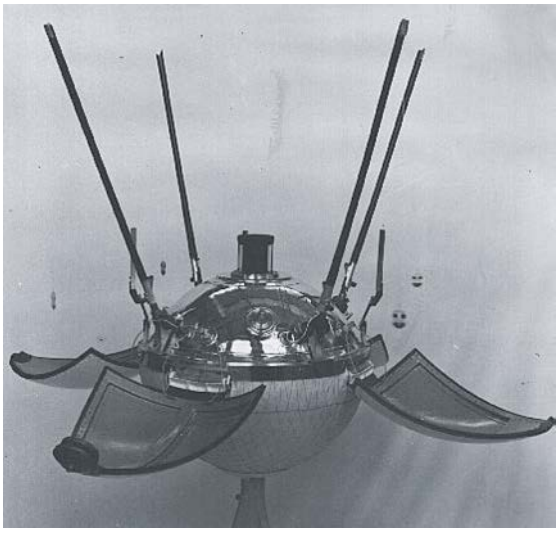


Figure 7.4.7: *Luna 9 with unfolded panels (Source: Lavochkin/ NASA).*

probably Pioneer 0 (a planned orbiter mission by NASA) and a Soviet Luna probe (probably planned as an impactor), both of which failed on launch in 1958.

Eventually, in March 1959, **Pioneer 4**, weighing only about 6 kg, was launched successfully and flew by the Moon. Radiation data was transmitted back to Earth.

The first probe which impacted on the Moon was **Lunik 2**, launched in September 1959.

In that same year, **Lunik 3** marked a great triumph for the Soviet space program by obtaining the first images from the far side of the Moon. A film was photochemically exposed, developed on-board the probe, fixed, scanned and the data then transmitted to Earth. Figure 7.4.6 shows an image of Lunik 3.

Luna 9 (Figure 7.4.7) achieved the first **soft landing** on the Moon in February 1966. The 99 kg probe transmitted panoramic images of the landing area.

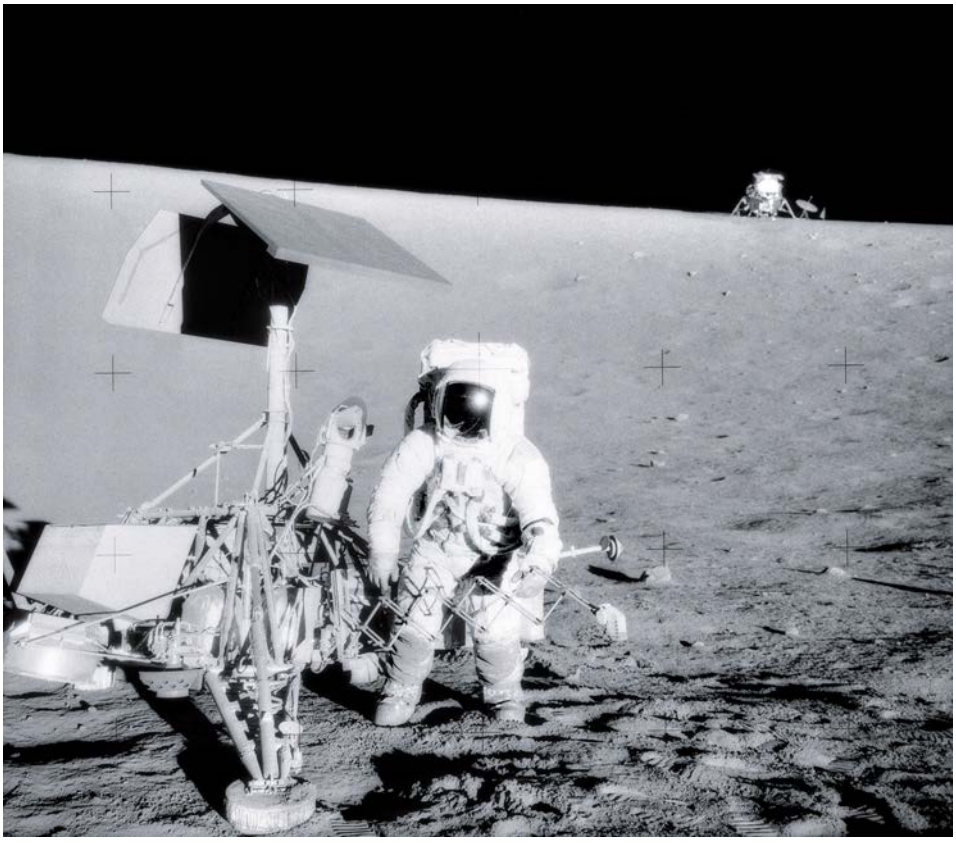


Figure 7.4.8: *Astronaut Conrad near Surveyor 3 (Source: NASA).*

Table 7.4.1: *List of Apollo Moon missions.*

	Apollo 8	Apollo 10	Apollo 11	Apollo 12	Apollo 13	Apollo 14	Apollo 15	Apollo 16	Apollo 17
Launch	21.12.68	18.05.68	16.07.69	14.11.69	11.04.70	31.01.71	26.07.71	16.04.72	07.12.72
Time spent on the Moon			21 h 36 min	31 h 31 min		33 h 30 min	66 h 55 min	71 h 2 min	75 h
Time of EVAs			2 h 31 min 1 EVA	7 h 46 min 2 EVAs		9 h 21 min 2 EVAs	18 h 33 min 3 EVAs	20 h 17 min 3 EVAs	22 h 4 min 3 EVAs
Returned lunar material			22 kg	34 kg		43 kg	77 kg	98 kg	117 kg
Landing area			Mare Tranquillitatis	Oceanus Procellarum		Fra Mauro	Hadley Rille	Descartes	Taurus-Littrow
Comments	First manned lunar mission	Test of lunar landing	First manned lunar landing		Mission aborted				Last Apollo lunar mission

NASA used three types of robotic spacecraft to prepare for the **manned Apollo missions**:

- **Ranger** (1961–1965): These impactors transmitted data back to Earth until shortly before impact. Ranger 7, 8 and 9 were successful; images were obtained with a resolution of up to 0.5 m.
- **Lunar orbiter** (1966–1967): A total of five successful orbiter missions were undertaken during which 99% of the lunar surface was mapped at a resolution better than 60 m.
- **Surveyor** (1966–1968): These were lander missions in preparation of Apollo. Figure 7.4.8 shows astronaut Charles Conrad inspecting Surveyor 3. Apollo 12 landed only about 180 m away from the landing site of the Surveyor.

Regarding further robotic missions, **Luna 16, 20 and 24** (1970, 1972 and 1976) should be mentioned; they successfully returned lunar samples back to Earth.

Luna 17 and 21 included the lunar rovers **Lunokhod 1 and 2** (1970 and 1973). The rovers were controlled in real time from the Earth using on-board cameras. An elaborate thermal control system based on radioactive polonium heater elements allowed Lunokhod to survive the lunar night.

More recent lunar missions have included only orbiters: Muses-A (Japan, 1990), Clementine (NASA, 1994) Lunar Prospector (NASA, 1998) and SMART-1 (ESA, 2003) Kaguya (Japan, 2007), Chang'e 1 (China, 2007), Chandrayaan 1 (India, 2008). Numerous further lunar missions (including landers) are being planned in the context of the worldwide Space Exploration Initiative.

Apollo

The highlight of lunar exploration is represented by the manned Apollo missions. A total of six landings (Apollo 11, 12, 14, 15, 16 and 17) and three manned lunar fly-bys (Apollo 8, 10 and 13) took place between 1968 and 1972 (Table 7.4.1).

The Apollo space vehicles, launched with Saturn rockets, consisted of a command module with places for three astronauts, a service module and the lunar module, the actual landing module (Figures 7.4.9 and 7.4.10). The three components together had a mass of about 50 tons.

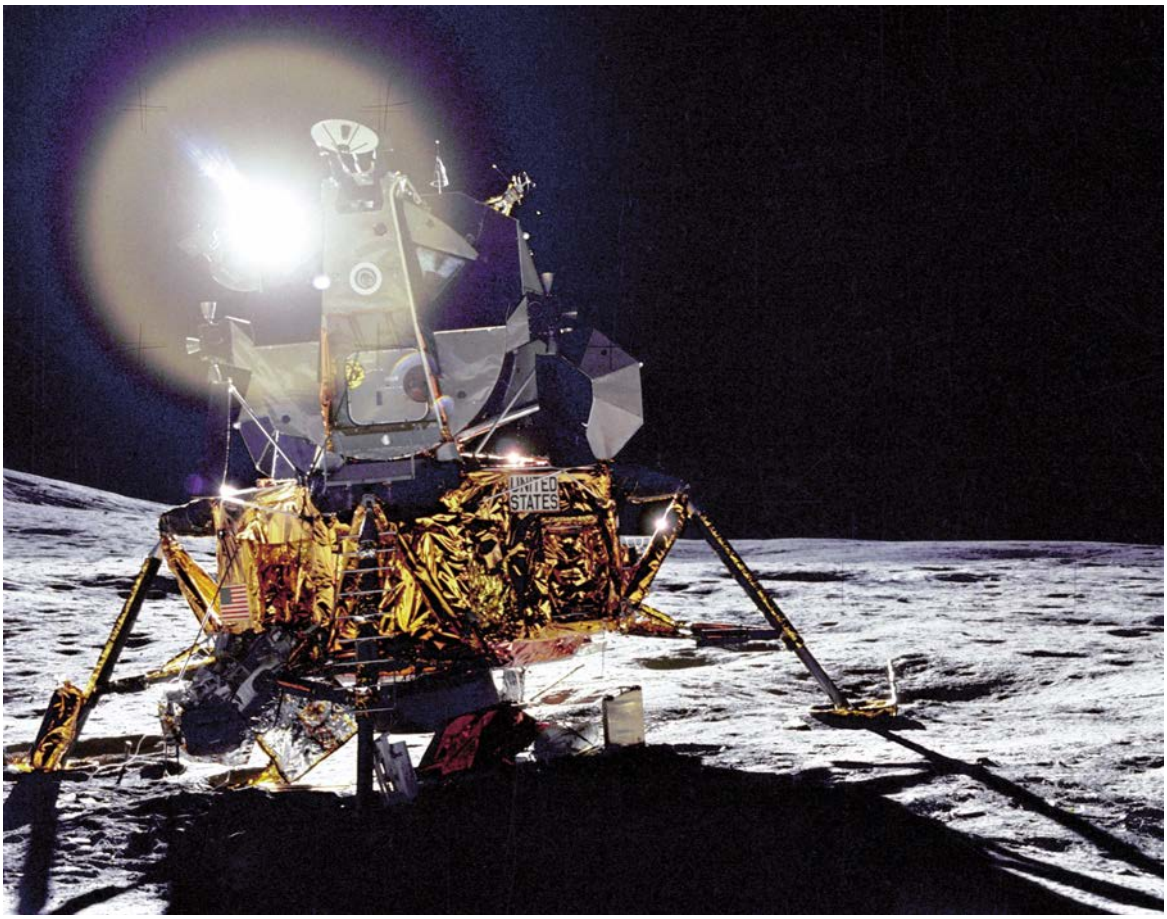


Figure 7.4.9: *The lunar module of Apollo 14 in Fra Mauro (Source: NASA).*

7.4.3 Planetary Missions

Very early in the space programs it was considered to send space probes not only to the Moon but also to Mars and Venus, which is a considerable challenge, bearing in mind the great distances involved. The technological aspects of interplanetary missions are addressed in more detail in Section 7.4.5.

After initial failures, the first successful mission to another planet took place in December 1962 with the **fly-by of Venus** by **Mariner 2** (NASA). The Mariner data revealed the retrograde rotation of Venus, and the high temperature and pressure as well as the composition (primarily CO_2) of the atmosphere. Mariner 2 (Figure 7.4.11) also discovered the lack of an intrinsic magnetic field.

Mariner 4, designed on the basis of the Mariner 2 bus, but with larger solar panels, transmitted the first data from a **Mars fly-by** in 1965. Table 7.4.2 lists selected missions to planets and small bodies in the Solar System.

Some of the spacecraft listed in Table 7.4.2 (partly due to their historical significance) are described below in more detail.

Viking

The two identical Viking probes each consisted of an orbiter and a lander. The scientific objective was



Figure 7.4.10: *The Apollo 15 service and command modules (Source: NASA).*

primarily **mapping the Martian surface**, determining the composition and physical characteristics of the **atmosphere** and the **Martian soil**, as well as searching for **indications of life**.

The wet mass of these probes (including orbiter, lander and fuel) was about 3500 kg. The dry mass of the orbiter was 900 kg, that of the lander 600 kg.



Figure 7.4.11: *Artist's impression of Mariner 2 (Source: NASA).*

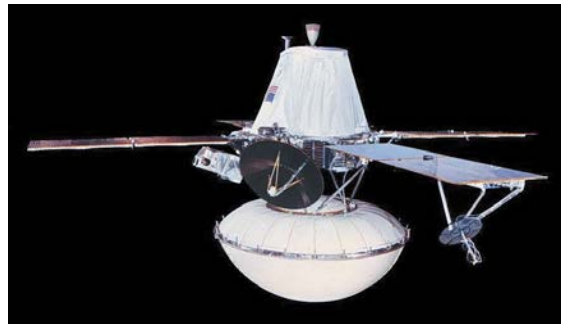


Figure 7.4.12: *Viking in cruise configuration; the lander is in the lower white container (Source: NASA).*

The **orbiters** were placed in an elliptical orbit (with a hydrazine/N_2O_4 engine of 1300 N thrust). They were three-axis-stabilized spacecraft which generated their electrical power from four solar panels (15 m^2; about 620 W in Mars orbit). Figure 7.4.12 shows Viking in cruise configuration.

The **landers** consisted of an aluminum structure and stood on three landing legs. They were decelerated before touchdown (after parachute separation) with hydrazine retrorocket engines with an adjustable thrust between 280 and 2670 N.

Power was generated by two **radiothermal generators** (RTGs) based on plutonium-238, each of which supplied 30 W (4.4 V) of electrical power. To cope with higher power demands, rechargeable NiCd batteries were also on-board.



Figure 7.4.13: *A model of the Viking lander (Source: NASA).*

Table 7.4.2: *Selected missions to investigate the Solar System.*

	Mission type	Launch date	Comments
Mariner 2	Venus fly-by	Aug. 1962	First successful mission to another planet
Mariner 4	Mars fly-by	Nov. 1964	First successful Mars fly-by
Venera 4	Venus atmospheric probe	June 1967	First atmospheric capsule on Venus, operational down to 25 km altitude
Venera 7	Venus lander	Aug. 1970	First planetary lander, transmitted until 23 min after landing
Mariner 9	Mars orbiter	May 1971	First Mars orbiter, approx. 1 year in operation, global mapping
Pioneer 10	Jupiter fly-by	Mar. 1972	First Jupiter fly-by, signals from the probe were received until 2003
Pioneer 11	Jupiter fly-by, Saturn fly-by	April 1973	Sister probe of Pioneer 10, 1979 first Saturn fly-by
Mariner 10	Mercury fly-by	Nov. 1973	1 Venus and 3 Mercury fly-bys
Helios 1 **Helios 2**	Heliocentric orbiter	Nov. 1974/ Jan. 1976	In-situ data from solar vicinity (< 0.3 AU)
Viking 1 **Viking 2**	Mars lander and orbiter	Aug. 1975/ Sept. 1975	Extremely successful Mars mission, years of operation and thousands of images, search for signs of life
Venera 9 **Venera 10**	Venus lander and orbiter	June 1975/ June 1975	Transmitted photographs of the Venusian surface
Voyager 2	Fly-bys of Jupiter, Saturn, Uranus and Neptune	Aug. 1977	Extremely successful mission, important for our understanding of the giant planets and the outer Solar System
Voyager 1	Fly-bys of Jupiter and Saturn	Sept. 1977	
Vega 1 **Vega 2**	Fly-by of Halley's Comet/Venus lander and balloon	Dec. 1984/ Dec. 1984	First images of a comet's nucleus, successful Venus landers and balloons
Giotto	Fly-by of Halley's Comet	July 1985	Fly-by close to Halley's Comet (and Grigg–Skjellerup)
Magellan	Venus orbiter	May 1989	Radar map of Venus
Galileo	Jupiter orbiter, atmospheric probes	Oct. 1989	First Jupiter orbiter, atmospheric entry probe, extremely successful
Ulysses	Heliospheric orbiter	Oct. 1990	Orbit passed over the solar poles
SOHO	Solar observation probe	Dec. 1995	Observes the Sun from the Lagrange point, still in operation
Cassini-Huygens	Saturn orbiter/Titan atmospheric probe/lander	Oct. 1997	Saturn orbiter/atmospheric probe to Saturn's moon Titan, transmitted images after landing
Deep Space 1	Asteroid and comet fly-by	Oct. 1998	Fly-by of Braille and Borelly, successful use of ion thrusters
Stardust	Comet fly-by and return capsule	Feb. 1999	Comet dust from the coma returned to Earth (Comet 81P/Wild 2)
Hayabusa	Asteroid sample return	May 2003	Investigation of Itokawa
Mars Express	Mars orbiter	June 2003	Improved mapping, remote sensing chemical analysis
Spirit (MER-A) **Opportunity (MER-B)**	Mars rover	June 2003/ July 2003	Extremely successful rover mission on Mars, still operational
Rosetta	Comet orbiter and lander (in flight)	Mar. 2004	Reaches Comet 67P/Churyumov–Gerasimenko in 2014
Messenger	Mercury orbiter	Aug. 2004	Mercury fly-bys, orbit insertion 2011
Deep Impact	Comet fly-by probe and impactor	Jan. 2005	Impact on Comet 9P/Temple, investigation of ejected material
MRO	Mars orbiter	Aug. 2005	Mars reconnaissance orbiter
Venus Express	Venus orbiter	Nov. 2005	Venus investigations
New Horizons	Pluto fly-by (in flight)	Jan. 2006	To reach Pluto in 2015, possible fly-bys of other Kuiper Belt objects



Figure 7.4.14: *Image of the surface of Mars taken by Viking 1 (Source: NASA).*

Figure 7.4.13 shows a model of the lander, and Figure 7.4.14 an image of the Martian surface (Chryse Planitia) as seen from the Viking 1 lander.

Communications with Earth were directly via an S-band or a UHF transponder system.

Each lander **payload** consisted of a set of 12 scientific instruments with a total mass of 91 kg. The Viking 2 lander ceased operation in April 1980, Viking 1 in November 1982; a total of about 1400 images were transmitted, in addition to meteorological data and chemical analyses of the Martian surface material.

7

Voyager

As an example of a mission to the outer Solar System, thus traveling far from the Sun, the Voyager missions will be briefly described.

The Voyager project was a huge success. Our understanding of Jupiter, Saturn, Uranus and Neptune (the latter two not being visited by any spacecraft since the Voyager 2 fly-bys in 1986 and 1989) was significantly influenced by the Voyager results.

Many of the technologies used for Voyager were also applied for larger probes such as Galileo (a Jupiter orbiter) and Cassini (a Saturn orbiter).

Voyager 1 and 2 were identical sister probes whose primary scientific goals were to investigate **Jupiter** and

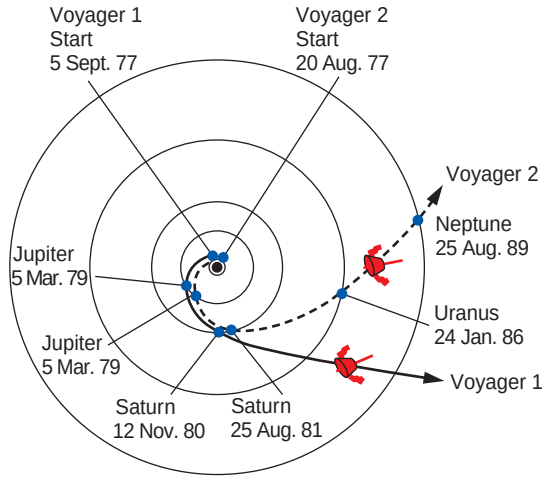


Figure 7.4.15: *Trajectories of the two Voyager probes (Source: NASA).*

Saturn as planetary systems (including atmosphere, magnetosphere, moons and ring structures). Because of the particular configuration of the planets, after the Saturn fly-by, Voyager 2 was directed to **Uranus** and further to **Neptune** (Figure 7.4.15).

The **primary structure** of each probe was a decangular (10-angle) bus (47 cm high, diameter: 178 cm). A parabolic high-gain antenna of 3.66 m diameter, an S-band system (uplink) and an X-band system (downlink, 160 bit/s) were used for communications with Earth.

Because of the great distance to the Sun (in the case of Saturn solar illumination is only 1/100th of that at Earth), solar cells could not be used for power generation. **Power** was obtained from three RTGs based on plutonium-238. At the beginning of the mission they delivered 470 W (30 V). In 2006 the delivered power was still 290 W (the half-life of plutonium-238 is about 88 years). Figure 7.4.16 is an artist's impression of the Voyager space probe.

In 2008, Voyager 1 was at a distance of about 107 AU (about $15 \cdot 10^7$ km) and Voyager 2 about 86 AU from the Sun; both are partially functional and continue to transmit data to Earth.

In December 2004 Voyager 1 detected the **termination shock** (an indication of the heliopause, the boundary between the Sun's sphere of influence and the interstellar medium).

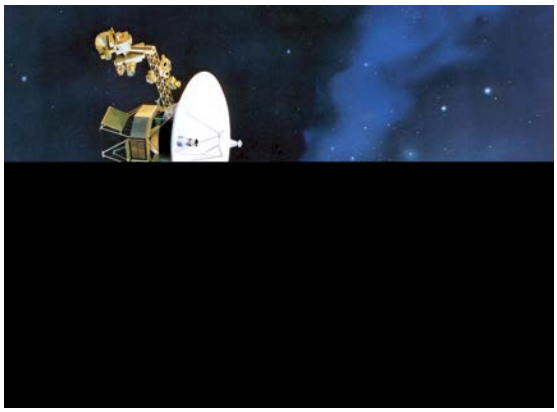


Figure 7.4.16: *Voyager (Source: NASA).*

Venera (7 to 14)

The Soviet Venera probes to **Venus** consisted of orbiters and landers. Venera 7 achieved the first soft landing on Venus in December 1970. Venera 9 and 10 landed in October 1975 and provided the first images of the planet's surface. A **cooling system** made it possible to continue to transmit data to Earth up to 53 (and 65, respectively) minutes after landing (the temperature of the atmosphere at the landing site was measured as 485 °C at a pressure of 90 bar). Figure 7.4.17 shows the Venera 9 lander.

The Venera 9 and 10 orbiters were used for data relay for the landers as well as for remote observations of the atmosphere.

The launch mass of the orbiter and lander together was about 5 tonnes in each case; the mass of the probe at entry (including the heat shield) was 1560 kg and the mass of the actual landing unit 660 kg.

Table 7.4.3 provides some information on the Venera lander missions.

Vega 1 and 2 **had fly-bys of Comet Halley and continued their mission** to Venus. Besides the orbiters and landers, two balloons were included.

Helios

Although not a planetary mission according to the definition, Helios is interesting from a technological point of view because of the challenging **thermal design.** Helios 1 and 2 were developed and assembled in Germany and launched together with a German–US payload from Cape Canaveral. The two spacecraft

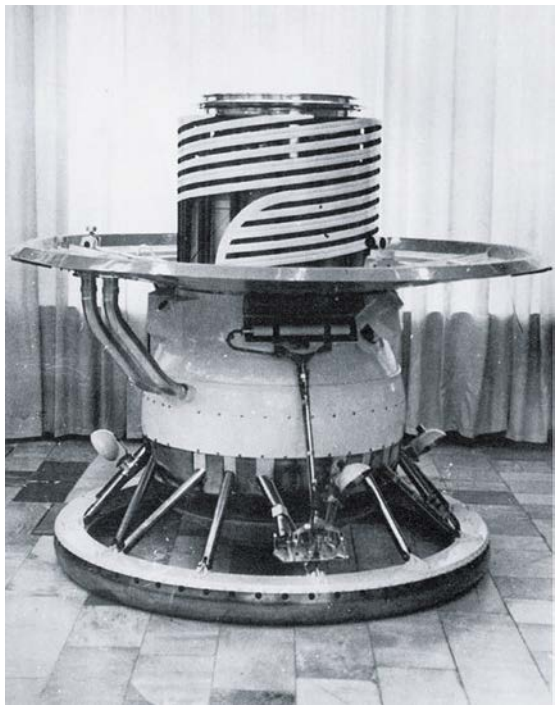


Figure 7.4.17: *Venera 9 lander (Source: Lavochkin/NASA).*

were placed in a heliocentric orbit within Mercury's orbit (0.29 AU, < 45 million km).

Helios 1 and 2 were identical sister probes. Both were spin stabilized. The characteristic narrow-waist, spool-type shape was chosen in order to obtain great radiative emission compared to the incoming radiation.

Table 7.4.3: *Venera lander missions.*

	Year	Launch mass [kg]	Operating time on the Venusian surface
Venera 7	1970	1180	23 min
Venera 8	1972	1184	50 min
Venera 9	1975	4936	53 min
Venera 10	1975	5033	63 min
Venera 11	1978	4540	95 min
Venera 12	1978	4540	110 min
Venera 13	1981	4363	127 min
Venera 14	1981	4715	57 min
Vega 1	1984	4950	56 min
Vega 2	1984	4950	57 min

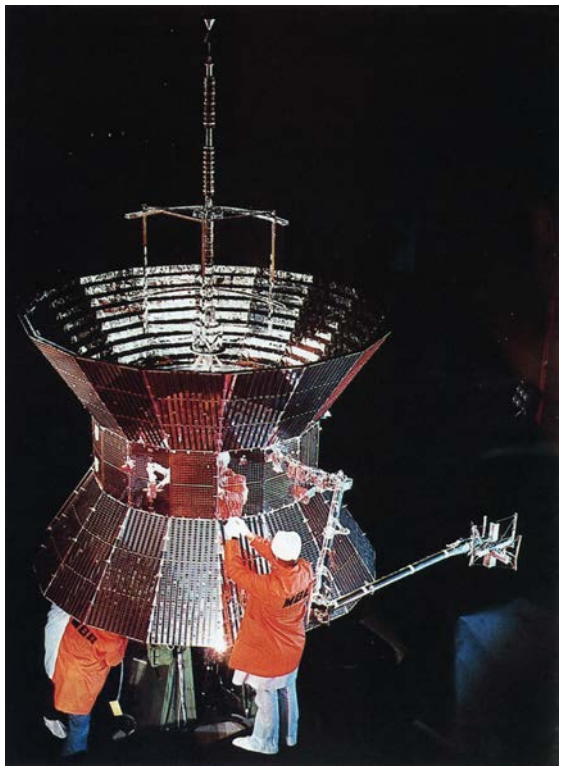


Figure 7.4.18: *Helios (Source: MBB).*

The sides were alternately covered with solar cells and second surface mirrors (SSMs). Figure 7.4.18 shows Helios during test preparations.

Both probes functioned for many years and provided extremely valuable data for solar physics.

7

Rosetta

Rosetta is an ESA mission planned to investigate in detail the **nucleus and environment of an active comet** (Comet 67P/Churyumov–Gerasimenko). Since comets are believed to consist of pristine material, almost unchanged since the time of the formation of the Solar System, their investigation is a key to understanding the history of our planetary system. Comets are also interesting in respect to the origin of life, since they may have transported organic material to Earth.

Rosetta was launched in March 2004 and is now in its cruise phase. The target comet will be reached in 2014



Figure 7.4.19: *Artist's impression of Rosetta during the Mars swing-by maneuver (Source: ESA).*

after a total of three **Earth swing-bys** and one **Mars swing-by maneuver**. The long flight phase is one of the technical challenges of the mission (Figure 7.4.19).

With the help of a navigation camera the comet can be precisely located upon approach, so that accurate orbit insertion will be possible. The **rendezvous** will take place at a heliocentric distance of over 4 AU. Rosetta is designed without radioactive heater units, and thus requires very large solar panels (a total of 64 m^2) and special solar cells which are highly efficient at low temperatures and low solar illumination.

The probe weighs about 3 tonnes (about half of which is fuel) and carries a total of 11 scientific experiments on-board to investigate the comet from orbit. The mission also includes a **lander** (***Philae***), which after remote characterization of the nucleus and selection of a suitable landing site will descend to the surface of the comet at about 3 AU from the Sun. *Philae*, with a total mass of 97 kg, will conduct in-situ investigations on the nucleus. *Philae* has a total of 10 scientific instruments on-board, including mass spectrometers, gas chromatographs and cameras (Figure 7.4.20).

It is important to securely anchor the lander onto the comet because its gravitation is so low that activities like drilling would not be possible otherwise. Harpoons with a tether will be used for the purpose; they will be fired immediately after touchdown.

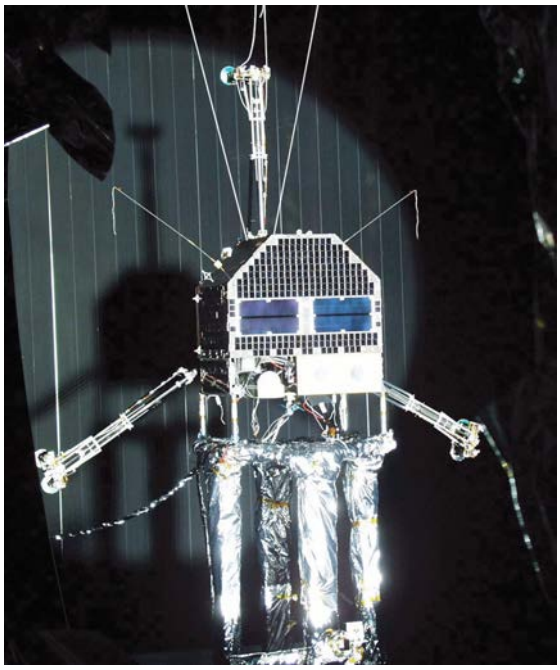


Figure 7.4.20: *The Rosetta lander Philae during thermal tests (Source: DLR/ESA).*

7.4.4 Mission Analysis of Interplanetary Space Probes

Since mission planning for space probes which leave Earth orbit requires considerations beyond those for Earth satellites, some of the aspects of orbit dynamics for interplanetary probes are discussed below.

7.4.4.1 Interplanetary Transfer

In order to inject a spacecraft into an interplanetary orbit, obviously, Earth's gravity field must be overcome, so the probe has to be accelerated to a velocity higher than the **escape velocity**. Assuming as starting point a circular parking orbit at about 320 km altitude (e.g., the scenario for Mars Express, launched with Soyuz Fregat; orbit velocity 7.7 km/s), an acceleration of an additional 3.2 km/s is required, as a minimum, in order to reach the escape velocity of 10.9 km/s.

To reach another planet, the probe must be put into an orbit around the Sun where (in the energetically optimum case) **aphelion** and **perihelion** are at Earth and at the target body, respectively. Such a **transfer orbit**, in which launch and arrival are exactly opposite to the Sun, is called **Hohmann transfer**.

Table 7.4.4: *Velocities for Hohmann transfer orbits.*

	Launch [km/s]	Arrival [km/s]	Flight time [a]
Mercury	−7.5	9.6	0.3
Venus	−2.5	2.7	0.4
Mars	2.9	−2.7	0.7
Jupiter	8.8	−5.4	2.7
Saturn	9.6	−5.4	6.0

Table 7.4.4 lists the velocities required for trajectories from Earth to planets, as well as the velocity relative to the planet at arrival. The second column gives the change in velocity required to leave the orbit of the Earth around the Sun and reach the orbit of a particular planet by making use of a Hohmann transfer (negative numbers indicate a deceleration). The third column gives the relative velocities at which the target planet is reached, and the last column the transfer time in years.

Since, as mentioned, start and arrival are 180° apart in the solar orbit, the Earth and the target must be in the correct position to each other at the time of launch. Accordingly, there are only specific **launch windows** available for such missions. There are good launch possibilities for flights from Earth to Mars every 2.13 years and from Earth to Venus every 1.6 years. (For the sake of completeness one should also mention the possibility of a 540° transfer. This opens more launch windows at the expense of a longer flight time.) Because of the inclination and eccentricity of planetary orbits, not all launch windows are energetically equally attractive.

7.4.4.2 Orbit and Rendezvous Missions

If on arrival the probe is to be put into an orbit around the target, then the **arrival velocity** (see Table 7.4.4) has to be reduced. The **velocity change** required, Δv, is shown in Equation 7.4.1 (with the escape velocity v_e and the arrival velocity v_a):

$$\Delta v = (v_e^2 + v_a^2)^{0,5} - v_e \qquad (7.4.1)$$

It is therefore desirable to carry out this maneuver as close as possible to the target. For example, the orbit insertion velocity for Mars at 300 km altitude is 680 m/s:

$$\Delta v = (4.81^2 + 2.65)^{0,5} - 4.81 = 0.68.$$

After entry, some kind of **orbit adjustment** is usually necessary, such as circularization. As discussed in more detail in Section 7.4.5, deceleration can be supported by aerobreaking (gentle breaking by atmospheric friction).

7.4.4.3 Swing-by Maneuvers

In order to reduce the necessary velocity change to be carried out with engines on interplanetary missions, so-called swing-by (**gravity-assist**) maneuvers can be performed.

When the probe flies through the **gravity field** of a planet, part of the **momentum** of this planet around the Sun is transferred to the spacecraft. If the planet is moving away from the probe on its orbit around the Sun, the probe gains kinetic energy (and vice versa).

Without gravity assist, many interplanetary missions would not be possible. In the case of the Rosetta mission, for example, four swing-by maneuvers are required (Earth, Mars, Earth, Earth). The resulting velocity change adds up to a total of about 19 km/s.

If one or more swing-by maneuvers are required for a mission, there are fewer **launch windows**, since the planets involved have to be in the right relative configuration. For example, a mission to Saturn with a swing-by at Jupiter is only possible about every 20 years. The swing-by situation used for Voyager 2 (Jupiter, Saturn, Uranus and Neptune) occurs only every 176 years!

7.4.5 Key Technologies for Planetary Missions

In conclusion, some of the key technologies which are particularly important for interplanetary missions are described below.

7.4.5.1 Thermal Control Systems for Extreme Environmental Conditions

Since, by definition, missions in the Solar System take place over a wide range of heliocentric distances, **thermal design** is often a particular challenge.

Usually, technologies which are also used for satellites in Earth orbit are used (insulation, radiators, heat pipes, a cleverly chosen relationship between absorption and emission (α/ε) of the surfaces involved), but are adapted for the thermal conditions of the particular mission.

Missions which move far away from the Sun or have to survive long periods without solar illumination (lunar night) are usually dependent on the use of **radioactive heater units** (RHUs). Radioactive decay produces heat: 1 g of plutonium-238 provides about 0.5 W of thermal power. Together with a **thermoelectric generator** (TEG), electrical power can also be obtained based on the Seebeck effect, which can be utilized for **radiothermal generators** (RTGs).

Because of the high toxicity of plutonium, the launch of probes containing this material is not only demanding with respect to the required safety measures, but also politically problematic.

High demands are made on the **thermal insulation** of interplanetary probes. Usually, **multilayer insulation** (MLI) is used. It consists of numerous (typically about 10) layers of reflective plastic foils (e.g., aluminum on kapton), which functions (in a vacuum) similar to a "multiwalled thermos flask." Under atmospheric conditions (as in the case of Mars or Venus landers), other methods have to be applied, for example **aerogels**, which have a very low thermal conductivity.

If **high-energy radiation** causes problems (either directly from the Sun or as backscatter, e.g., from Venus) then coated mirrors (**second surface mirrors** (SSMs)) with a high reflection coefficient in the visible range and high emissivity in the infrared are used to cover parts of the surface of the spacecraft.

7.4.5.2 Radio Thermal Generators

As mentioned above, the heat generated by radioactive material can be used together with a radiator to produce electrical power by making use of the

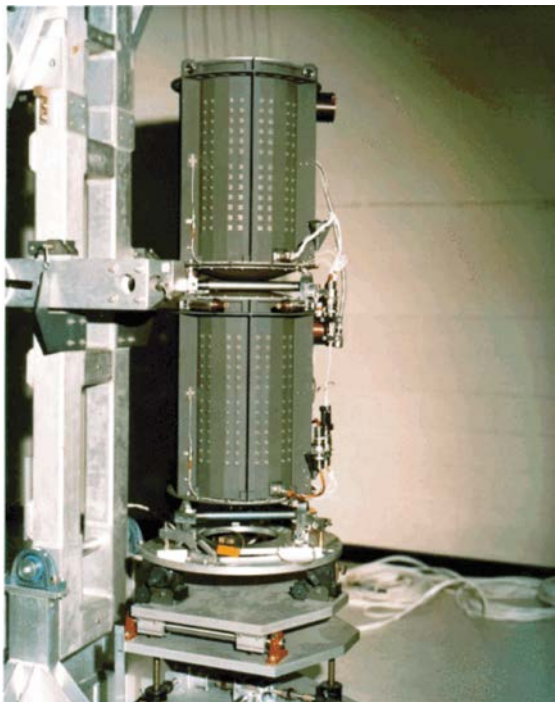


Figure 7.4.21: *Testing of the Voyager RTG (Source: NASA).*

Seebeck effect. Such systems are needed for missions beyond the orbit of Jupiter, where solar cells are no longer useful.

Several semiconductors (e.g., Bi_2Te_3) are suitable for producing electrical voltage if there are sufficiently high temperature gradients. An RTG therefore consists of a **heating element** (often plutonium-238 as oxide), a **thermoelectric generator** (TEG) and a **radiator** for the cold side.

RTGs were used, for instance, for Pioneer 10 and 11, Voyager, the Viking landers, Ulysses, Galileo, Cassini, and for long-term experiments placed on the Moon by Apollo. Figure 7.4.21 shows a Voyager RTG.

7.4.5.3 Landing Systems

One special category of planetary missions includes surface **landing**. In principle a distinction has to be made between landing on bodies without an atmosphere (like our Moon), on those with an atmosphere (Mars, Venus, Titan) and small bodies (asteroids and comets).

In the case of the last bodies, a low impact velocity can be chosen because of the low gravitational forces. High-impact energy and landing shock are thus not as problematic as for landings on planets or large moons. On the other hand, the calculation and navigation of the descent as well as the avoidance of any rebound are difficult.

The landing on a **body with an atmosphere** can be supported by braking with an entry shield and a parachute system. The particular scenario depends on the density and pressure of the atmosphere (the ground pressure on Mars is about 10 mbar, on Venus 90 bar and on Titan 1.6 bar), as well as on its chemical composition and the scale height.

Landing on large bodies without an atmosphere requires **braking rockets**, a fact which always leads to an unfavorable ratio between spacecraft mass and payload. This is the case for landing on the Moon, but also, for example, on Mercury or Europa (there have been concrete plans to do so).

The **landing** itself either is **controlled** and typically carried out on either landing legs (e.g., for Surveyor and Viking) or a landing platform (as for Venera), or involves an **airbag system** which encases the entire lander when it hits the ground. In the latter case the probe usually has to right itself (as for Luna 9, Mars Pathfinder, MER) after the airbag is released (Figure 7.4.22).

For controlled landings, **engines that can be throttled** and **altimeters** are typically used.

7.4.5.4 Long-Distance Communication Systems

Since interplanetary probes are much further away from the Earth than are satellites, communication, consequently, is more demanding.

A decisive parameter in the operation of such missions is the **signal propagation time** which is required for a light or radio wave to travel from the Earth to the probe (and vice versa). The speed of light is about $3 \cdot 10^8$ m/s, which means that the propagation delay to a probe at a distance of 1 AU is 8 min. From Saturn the signal propagation time is already up to 1.5 h.

Commands (as well as data), thus, arrive only after considerable delay, which makes a high degree of **autonomy** and automatic failure detection and recovery necessary.

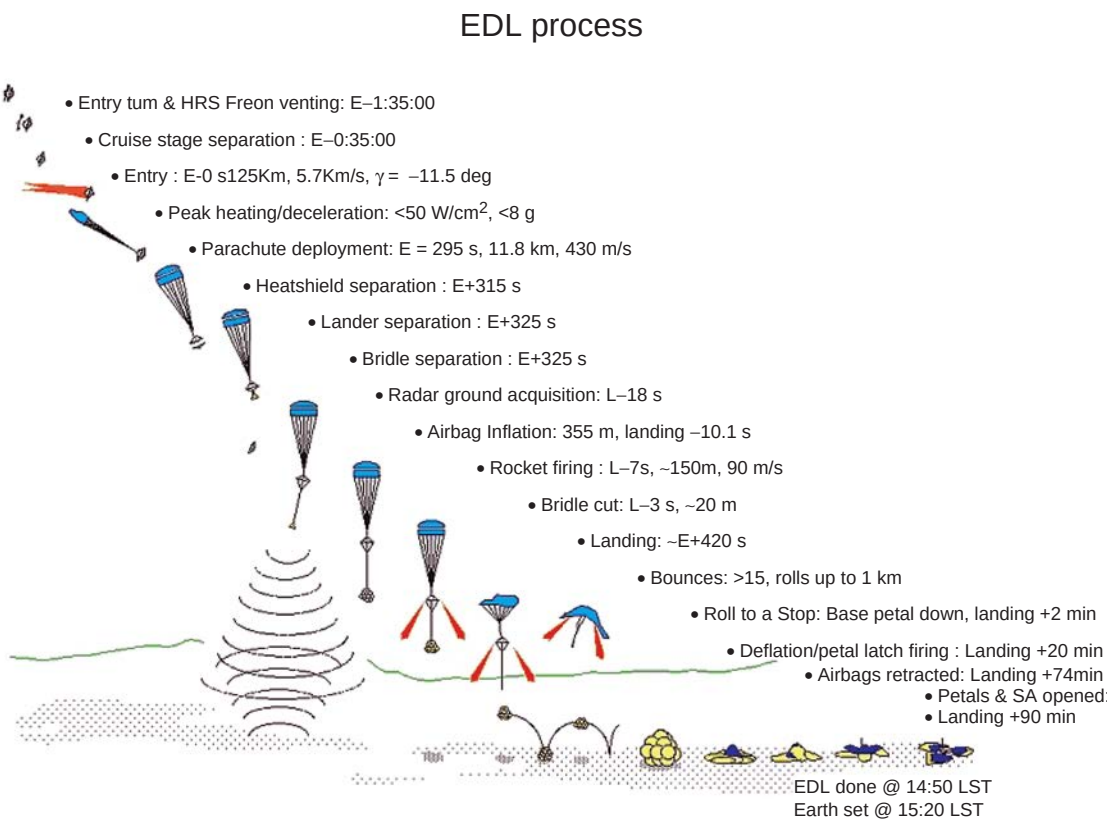


Figure 7.4.22: *Entry, descent and landing scenario for the MER lander (Source: NASA).*

In order to communicate over such distances, compromises have to be made concerning the **data rate**, or relatively **large parabolic antennas** must be installed on the probe and large and sensitive ground stations are needed on Earth. ESA uses for communications with interplanetary probes the 35 m antennas in New Norcia, Australia, and Cebreros, Spain (see Figure 7.4.23), NASA the Deep Space Network (DSN) with 35 m and 70 m ground antennas.

For communications, the **X-band** is normally used (frequently with a backup system in the **S-band**). In the future, technologies also using the Ka-band will be more common (Table 7.4.5).

The space probes have both **high-gain antennas** for data transmission as well as antennas with lower directivity so that contact can also be established if there is bad pointing at the Earth (or even uncontrolled spin nutation). This can save the mission in emergency situations.

Planetary landers frequently use an orbiter as relay (e.g., the nominal case for the Mars Exploration rovers, or in the case of the Rosetta lander from the comet surface).

7.4.5.5 Navigation

The navigation of a space vehicle involves two aspects:

- Determining the position and velocity vector.
- **Controlling** the probe by firing the engines.

Table 7.4.5: *The most common frequency bands for communications with space probes.*

Band	Frequency range
S-Band	2.7 – 3.5 GHz
X-Band	8 – 12.5 GHz
Ka-Band	26.5 – 40 GHz



Figure 7.4.23: *ESA ground station in Cebreros, Spain (35 m diameter antenna) (Source: ESA).*

As with communications, the same methods can be used for interplanetary missions as for Earth-orbiting satellites, but the large distances involved make the task more challenging.

So-called **orbit insertion maneuvers** need to be mentioned, where a space probe is placed into orbit around a planet upon approaching it (see Section 7.4.4). In order to save fuel to reach the desired orbit, the friction of the upper atmosphere can be used for smooth deceleration (aerobraking, as demonstrated with the Mars Global Surveyor).

Bibliography

[7.4.1] Lindberg Christensen, L., Fosbury, R. *Hubble. 15 Jahre Entdeckungsreise*. Wiley-VCH Verlag GmbH, 2006.

[7.4.2] Schlegel, E. *The Restless Universe: Understanding X-ray Astronomy*. Oxford: Oxford University Press, 2002.

[7.4.3] Cesarsky, C., Salama, A. (eds.) ISO Science Legacy. A Compact Review of ISO Major Achievements. Reprinted from *Space Sci. Rev. J.*, **119** (1–4). Heidelberg: Springer Verlag, 2006.

[7.4.4] Information zu fast allen Astronomie- und Planetenmissionen. http://nssdc.gsfc.nasa.gov/.

[7.4.5] Internetseite zu den ESA Missionen. http://www.esa.int/esaSC/index.html.

[7.4.6] Internetseite mit Informationen zu russischen und sowjetischen Missionen. http://www.russianspaceweb.com/.

[7.4.7] Ball, A. *et al. Planetary Landers and Entry Probes*. Cambridge: Cambridge University Press, 2007.

[7.4.8] Scientific Results of the Viking Project. *J. Geophys. Res.*, Special Issue, **82** (28), 1977.

[7.4.9] Glassmaier, K.-H. *et al.* The Rosetta Mission: Flying towards the Origin of the Solar System. *Space Sci. Rev.*, Special Issue, **128**, 2007.

[7.4.10] Evans, B. *NASA's Voyager Missions: Exploring the Outer Solar System and Beyond*. Heidelberg: Springer Verlag, 2003.

[7.4.11] Stone, E. *et al.* Voyager 1 Explores the Termination Shock Region and the Heliosheath Beyond. *Science*, **309** (5743), 2017, 2005.

[7.4.12] Squyres, S. Roving Mars: Spirit, Opportunity, and the Exploration of the Red Planet. *Hyperion*, 2005.

[7.4.13] McFadden, L.A., Weissman, P., Johnson, T. (eds.) *Encyclopedia of the Solar System*, Second Edition. New York: Academic Press, 2006.

[7.4.14] Surkov, V. *Exploration of Terrestrial Planets from Spacecraft: Instrumentation, Investigation, Interpretation*, Second Edition. New York: John Wiley & Sons, Inc., 1997.

[7.4.15] Harland, D. *Exploring the Moon: The Apollo Expeditions*. Heidelberg: Springer Verlag, 1999.

7.5 Materials Science

Ivan Egry

The term **materials science** is used for basic and applied investigations of materials in their solid, liquid and fluid states.

The interest of materials scientists in using space for their own research is based almost exclusively on the opportunity to conduct experiments under reduced gravity conditions, **microgravity**, as it exists on spacecraft. Therefore, this section begins with a discussion of the origins and effects of microgravity, before turning to the more topical themes.

The section ends with a discussion of current or planned experimental equipment, **payloads** in the technical context of space activities, which can be used to conduct these experiments.

7.5.1 Microgravity

7.5.1.1 Origins

Gravity is one of the fundamental forces of the Universe. As with the electrostatic Coulomb force, it acts over long distances. Whereas electric fields exert their force only on charged (polarizable) bodies, gravity affects all bodies with mass. It owes its name to the fact that it interacts with the gravitational mass, in contrast to the inertial mass of a body which describes its resistance to the change of motion.

That gravity acts over long distances has the consequence that Earth's gravitational field extends far into space. The force of Earth's gravity at the altitude of the International Space Station (ISS) is, for example, still about 90% of its value on the surface. So the reason for weightlessness on the ISS is not because there is no gravity field there. The phenomenon of apparent weightlessness has a dynamic cause and is related to the system of reference from which one observes the movement of other bodies. If we were in an elevator whose support cables had broken, putting us in free fall, we could, assuming we were cool-minded enough, let some object in our possession drop and note that it was not moving relative to ourselves. For us it would be "weightless." But someone observing from outside would see that both we and the object were in free fall due to the force of Earth's gravity. In terms of classical physics, the outside observer is in an inertial frame of reference, whereas we are in an accelerated frame of reference compared to that inertial frame of reference. Newton's laws of classical mechanics are only true for an inertial frame of reference. For the laws of mechanics also to be valid for an accelerated frame of reference, we must postulate an additional force which cancels the effect of gravity so that our observation of the resting body is in accord with Newton's laws. This force is called the **fictitious force of inertia** F_f and is oppositely proportional to the acceleration of the frame of reference relative to the inertial frame of reference:

$$F_f = -m\ddot{s} \tag{7.5.1}$$

With the help of this fictitious force of inertia one can write the equation of motion of a point mass m in an accelerated frame of reference with the coordinates r' as follows:

$$m\ddot{r}' = m\ddot{r} + F_f \tag{7.5.2}$$

with r the coordinates of the point mass in the inertial frame of reference.

In the example of the free-falling elevator, $\ddot{r} = -g$ and $\ddot{s} = -g$. Inserting these into Equation 7.5.2 it follows that $m\ddot{r}' = 0$

Equation 7.5.2 is the basis for describing the phenomenon of weightlessness. It is true for the case of linear acceleration as in the ZARM drop tower in Bremen, for weightlessness during parabolic flight, and for space vehicles orbiting the Earth like the ISS. It should be mentioned that for the force of inertia it is the acceleration and not the velocity which is critical. So the phase of weightlessness during a parabolic flight already begins during the ascending portion of the parabola, and accordingly ZARM recently introduced an upward-catapulted launch instead of a regular drop, thereby doubling the duration of weightlessness.

In the case of rotational acceleration, the fictitious force of inertia takes the form of **centrifugal force**. What happens on an experiment platform in orbit can be most easily described by this force.

For a point mass m which circles the Earth at a distance R and angular velocity ω, the centrifugal force amounts to

$$F_c = m\,\omega^2\,R \tag{7.5.3}$$

This force has to compensate for the Earth's force of attraction, which is given by

$$F_g = -\frac{\gamma M\, m}{R^2} \tag{7.5.4}$$

where γ is the gravitational constant and M the mass of the Earth. By equating both forces one obtains

$$\omega^2\,R^3 = \gamma M = \text{const.} \tag{7.5.5}$$

which is nothing more than Kepler's third law. It says that the squares of the orbital periods ($T = 2\pi/\omega$) are directly proportional to the cubes of the radii.



Figure 7.5.1: *Isolines of residual gravity on the ISS, viewed opposite to the flight direction (Source: ESA).*

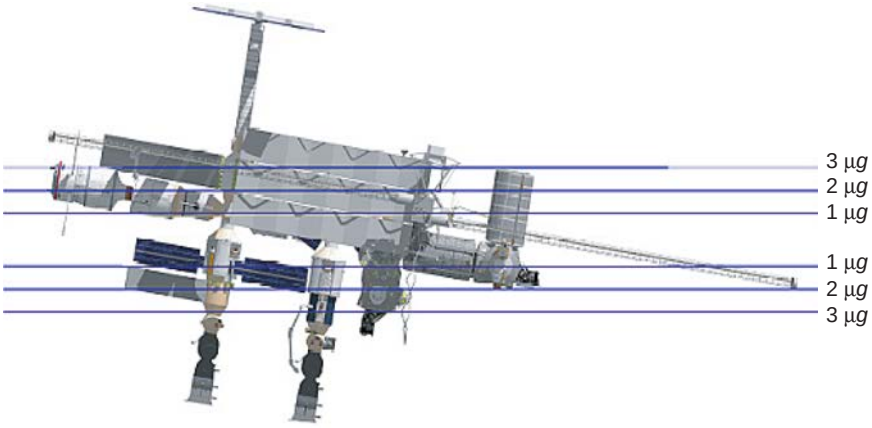


Figure 7.5.2: *Isolines of residual gravity on the ISS, viewed perpendicular to the flight direction (Source: ESA).*

Considering both Equations 7.5.3 and 7.5.4, it becomes clear that weightlessness can only be valid for a point mass. For an elongated body such as the ISS, these forces are compensated only along a line which approximately corresponds to the trajectory of the center of gravity of the space station (Figures 7.5.1 and 7.5.2).[2] If one moves away from the center of gravity in a plane perpendicular to the flight direction, compensation is not complete and one experiences a **residual force of gravity** or "microgravity." At an altitude of 300 km from the surface of the Earth this residual gravity amounts to about $0.3 \cdot 10^{-6} g_0$/m in the radial and $0.1 \cdot 10^{-6} g_0$/m in the lateral direction, where g_0 is the value of the acceleration due to gravity on the Earth's surface ($g_0 = 9.81\ \mathrm{m/s^2}$).

[2] For relatively compact bodies like the ISS, this is a very good approximation. For extended bodies like tethered satellites, a more detailed discussion is necessary.

In addition to the effect described above, there are other detrimental influences which inevitably lead to residual gravity. One of them is the air resistance or drag of the space vehicle, which at this altitude is not negligible. It depends on the density of the remaining atmosphere, the air stream over the surface and the coefficient of friction.

For the ISS this leads to a residual gravity on the order of $10^{-6} g_0$. There are dedicated experiment platforms which compensate for this effect with a cold gas propulsion system.

Besides these basically time-independent disturbances there are also **dynamic effects** which cause time-dependent residual gravity. In this category can be listed the firing of attitude control propulsion system, but also all moving objects, including the astronauts themselves, which, following Newton's law that for every action there is an equal and opposite reaction, cause the space vehicle to experience

an impulse change. These disturbances can reach an amplitude of $10^{-3} g_0$, sometimes even higher. It is accordingly important when constructing subsystems to keep these disturbances to a minimum. One can also attempt to isolate the experimental payload from the rest of the space vehicle by active damping. This approach is being taken in the ISS with the so-called ARIS system. Finally, one can cope with the unavoidable remaining disturbances by clever scheduling ("timelining") so that they do not unnecessarily disrupt the experiments.

7.5.1.2 Causes

After clarifying the origin of microgravity, it remains to discuss the question of why materials science experiments benefit from the microgravity environment and to what extent time-dependent and time-independent residual gravity can be tolerated.

First of all, it should be clear that the influence of gravity can be ignored for a solid body at rest, since its inner structure and outer form are primarily influenced by Coulomb forces. This is true independent of the form of atomic binding, for example covalent, ionic or metallic. The influence of gravity on gases is expressed, for example, in the **barometric altitude formula**. However, thermal fluctuations in gases are usually so strong that they conceal the influence of gravity over short distances (and for brief times). The liquid state is the most sensitive to gravity influences. This is because, in contrast to a solid, they have a **free surface** and cannot sustain **shear forces**. Therefore, most materials science experiments under microgravity conditions concentrate on the liquid and (to a lesser extent) gaseous states.

Under weightless conditions there is neither an up nor a down, neither heaviness nor lightness. A weightless drop of liquid remains in position and does not flow; it is therefore not necessary to store it in a container. In particular, weightlessness means that the following physical effects do not occur in the fluid states:[3]

- Buoyancy
- Sedimentation
- Convection.

[3] In thermodynamics one subsumes the liquid and gaseous states under the term **fluid state**.

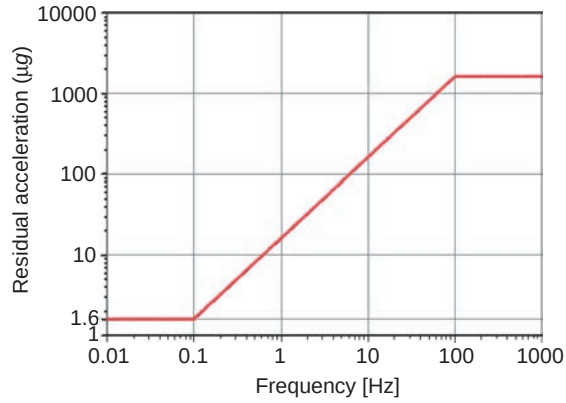


Figure 7.5.3: *Allowable microgravity disturbance on the ISS.*

By buoyancy is meant the upward motion of the lighter components in a liquid mixture, by sedimentation the downward motion of the heavy components. Convection, or more precisely buoyancy convection, refers to a closed fluid flow which is driven by the interaction of temperature and density gradients. (There are other kinds of convection that will be discussed later.)

The absence of these effects makes it possible to investigate phenomena which they normally conceal.

Just how sensitive individual experiments are to microgravity disturbances (Figure 7.5.3), or which level of microgravity they can tolerate, has to be decided from case to case. However, an estimate of the allowable amplitude of disturbances as a function of their frequency is possible with the help of the **Navier–Stokes equation**, which describes the macroscopic behavior of fluid systems in the context of hydrodynamics. Assuming permissible temperature variations of δT, then

$$g(\omega) = \frac{\delta T}{\nabla T} \frac{\rho / L}{\nabla \rho} \sqrt{\omega^2 + (\nu / L^2)^2} \sqrt{\omega^2 + (\kappa / L^2)^2} \qquad (7.5.6)$$

where:
ω = angular frequency of the disturbance,
∇T = temperature gradient,
$\nabla \rho$ = density gradient,
ν = kinematic viscosity,
κ = thermal diffusivity,
L = typical system length.

As rough as this approximation is, it nevertheless forms the foundation for the **specification for the ISS**, which is ($\omega = 2\pi f$)

$$g(f) = \begin{cases} 1 \cdot 6 \cdot 10^{-6} g_0 & 0,01 < f < 0,1 \text{ Hz} \\ f \cdot 16 \cdot 10^{-6} g_0 & 0,1 < f < 100 \text{ Hz} \\ 1600 \cdot 10^{-6} g_0 & 100 < f < 300 \text{ Hz} \end{cases} \quad (7.5.7)$$

Logically, experiments outside this specification should not be accommodated on the ISS. They require dedicated unmanned experiment platforms.

The various classes of materials science experiments which benefit from the microgravity environment are described in detail below. Some of them have already been carried out, others are planned.

7.5.2 Critical Phenomena

Matter can exist in various aggregate conditions and phases. By changing external thermodynamic variables such as pressure and temperature, one can transform such a system from one phase to another. This process is called a **phase transition**. Examples are the transition from a solid to a liquid state (melting), from the liquid to the gaseous state (evaporation), from an ordered to an unordered solid state, from a magnetic to a nonmagnetic state, etc. Ehrenfest classified phase transitions by the discontinuities in the thermodynamic variables which change at a phase transition. For a **first-order phase transition** it is the thermodynamic variables themselves, for example the density at the solid–liquid phase transition. In **second-order phase transitions** the variables themselves are continuous but their derivatives diverge. For example, the magnetization of ferromagnetic phase transitions changes continuously (from zero to a finite value), but the susceptibility, $\chi = \partial M / \partial H$, diverges as a function of temperature. Second-order phase transitions are characterized by a **critical point** in phase space. Only when all state variables are at their critical value is it a case of a second-order phase transition. For a fluid system the critical point is defined by the critical temperature T_c and the critical density ρ_c.

How individual thermodynamic values and their derivatives behave near a critical point can be described by the so-called **critical exponents**. If A is such a value and $\tau = (T - T_c)/T_c$, the dimensionless temperature, then the corresponding critical exponent α is defined as follows:

$$A = \tau^{\alpha} \Big|_{\tau \to 0} \quad (7.5.8)$$

The critical exponents do not just describe the behavior of individual values as they approach the critical point. There are also relationships between the individual critical exponents, so-called **scaling laws**. K. Wilson received the Nobel Prize in 1982 for confirming these scaling laws in the context of renormalization group theory.

The value of individual critical exponents is very difficult to calculate by theory. Different theories produce different values. It is also not clear how closely one must approach the critical point for Equation 7.5.8 to be valid, and what influence a finite system size has. Therefore, experimental determination of these exponents is extremely important. Under terrestrial conditions there is a difficulty with fluid systems: whereas the critical temperature can in principle be adjusted as precisely as desired, this is not the case for the density. Because of gravity the density at the bottom of the container will always be higher than near the top, and the critical density will only be found in an infinitely thin layer perpendicular to the force of gravity. The volume of fluid at the critical point is therefore minuscule compared to the total volume, and the anticipated critical behavior is masked by the noncritical behavior of the rest of the system. Here is a role for microgravity, since **density stratification** does not occur there and the entire volume can thus assume the critical value for density.

Experiments on critical phenomena have been conducted under microgravity conditions and more are being planned. Thus the specific heat $c_p = (\partial Q/\partial T)_p$ of SF_6 has been measured. The specific heat shows only a very weak (logarithmic) singularity; therefore its critical exponent is almost zero and hardly measurable with conventional methods.

In addition to classical systems, **quantum systems** are also of considerable interest. Thus the specific heat of ^4He along the λ line of the phase transition from the normal to the superfluid condition was measured with great precision (Figure 7.5.4). Recently it has

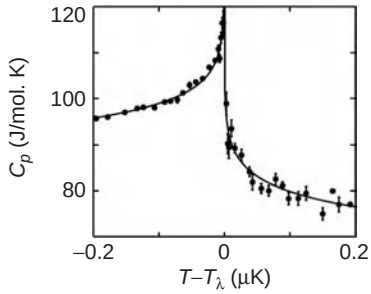


Figure 7.5.4: *Specific heat of ^{4}He near the λ transition (Source: J. Lipa, Stanford).*

been proposed that Bose–Einstein condensation be experimentally determined under microgravity conditions. This condensation is purely a quantum effect that only occurs with particles obeying Bose statistics[4] and it means that all subsystems condense in the quantum-mechanical ground state, and, in this way, manifest themselves macroscopically. Bose–Einstein condensation only occurs in the absence of external fields, including gravity, and can therefore be optimally observed under microgravity conditions for systems having mass.

7.5.3 Fluid Physics

7.5.3.1 Statics

Naturally enough, the **behavior of liquids** in a reduced gravity field is one of the central topics of materials science research under microgravity conditions. First of all there are questions about the **form of equilibrium** in liquid volumes and its stability. Here gravity is in competition with **surface tension**. While for the latter there is a tendency to minimize surface (relative to volume), leading to curved surfaces, in the volume of a liquid there is a tendency to minimize its potential energy in the gravity field, leading to the formation of flat surfaces as a response to the flat equipotential lines of terrestrial gravitation. The precise form of the surface is a compromise between these two conflicting principles and can be calculated using differential geometry. Qualitatively these effects can be discussed with the help of **capillary length**, which describes the relationship between surface energy and potential energy in a (constant) gravity field:

$$L_K = \sqrt{\frac{\gamma}{\rho\, g}} \tag{7.5.9}$$

where:
γ = surface tension,
ρ = density,
g = acceleration due to gravity.
If the capillary length is short relative to a typical system's dimension, then the influence of gravity dominates and one has flat surfaces. If it is large, then the influence of the surface tension is dominant and one has curved surfaces.

An interesting question concerns the **stability** of various liquid forms. The stability of a liquid bridge suspended between two concentric disks has been a matter of intensive investigation. Under weightless conditions, namely for infinite capillary length, the theory predicts the collapse of the liquid bridge as soon as its diameter is exceeded by the distance between the separating disks. This so-called Laplace instability can only be studied in its pure form under microgravity.

7.5.3.2 Dynamics

Besides these static effects there are dynamic phenomena which are the subject of hydrodynamic investigations under microgravity, including primarily **convection**. As already mentioned above, buoyancy convection is a result of the interaction of temperature gradients and gravity. Most bodies expand when they are heated; that is, they possess a positive thermal expansion coefficient. The consequence is that warmer liquids in a gravity field rise, whereas colder liquids sink. If a liquid is heated from below (as on a stove hotplate), flow or **buoyancy convection** occurs as soon as a critical value in the temperature gradient is exceeded. The geometry of the convection cell depends on the geometry of the entire system and changes with increasing temperature gradients until there

[4] Bose statistics are valid for particles with integer spin and allow several particles to exist in the same quantum state. This is to be contrasted with Fermi statistics, valid for particles with half-integer spin, in which case this is not possible because of the Pauli exclusion principle.

is a transition to turbulent flow. The characteristic quantity for describing buoyancy convection is the dimensionless **Rayleigh number**:

$$Ra = \frac{\beta \rho g |\nabla T|}{\kappa \eta} d^4 \qquad (7.5.10)$$

where:
β = thermal coefficient of expansion,
∇T = temperature gradient,
κ = temperature conductivity,
η = viscosity,
d = typical system length.
In the case of unstable layering, if the Rayleigh number exceeds a critical value Ra_c, then convection begins.

In the case of stable layering (heating from above) there is no convection. In the case of a radial temperature gradient (perpendicular to the force of gravity), then convection is initiated in all temperature gradients, no matter how small; the critical Rayleigh number is zero for this configuration. Because the Rayleigh number contains the value for the acceleration due to gravity, when it is changed then the threshold for the initiation of buoyancy convection accordingly changes, so the stability range of the instable layering can be considerably increased under microgravity.

Convection is a very efficient mechanism for transporting mass and heat in liquids. Only if it is suppressed is it possible to study other, weaker effects which are otherwise masked by the buoyancy convection. One of these effects is **Marangoni convection**, which was discovered at the beginning of the twentieth century but experienced a renaissance thanks to materials science experiments under weightlessness. In contrast to buoyancy convection it has its origin not in the temperature dependence of density, but in the temperature dependence of surface tension.[5] As a rule, surface tension increases as the temperature drops, therefore surfaces tend to arch more at cold locations than warm ones where the capillary length is smaller.

This leads to flow along the surface from warm to cold. Because the mass must be conserved, this flow has to be compensated within the liquid and convection cells are accordingly formed. Marangoni convection is independent of gravity. Since it has the same appearance as buoyancy convection, it can best be studied when the latter is suppressed by not exceeding the critical Rayleigh number, which can be accomplished by reducing gravity.

Besides liquids, other fluid systems have also been investigated under microgravity conditions. A particularly active field is **combustion research**. In an ordinary flame, but also during the combustion of fuel mixtures in motors, convection is the significant mechanism for mass and heat transport. **Diffusion processes** also play an important role, and these can only be studied if convection is suppressed. It was originally not known whether a flame would also burn under microgravity conditions; that is, whether the diffusion processes are sufficient for supplying the flame with oxygen and removing combustion products such as soot. Indeed, a candle burns under microgravity with a bluish, spherical flame.

Combustion research under microgravity conditions is not only undertaken to investigate aspects of basic and applied research to improve efficiencies; the safety aspects of fighting fires in a space vehicle, such as the ISS, also play a role.

7.5.3.3 Soft Matter

Among the fluid systems one also has to count other forms of soft matter like colloids, granular media, aerogels, aerosols, complex plasmas and cold atoms. They too are a subject of materials science research under microgravity conditions.

So-called **plasma crystals** are a spectacular result of this research. If one puts an electric charge on dust particles in a plasma at sufficiently low temperature, they align themselves in a regular crystal lattice, forming a plasma crystal. But on Earth these dust particles sediment; one can at best achieve a two-dimensional plasma crystal. Under microgravity conditions, however, the weight of the dust particles is irrelevant and a perfect three-dimensional crystal forms (Figure 7.5.5). In contrast to a crystalline solid which is composed of atoms, the length scale is larger by a factor of 10^4 (a dust particle is about 1 µm in diameter, an atom about 1 Å), which means that one can observe the equivalent of atomic processes with

[5] Or, more precisely, in the spatial dependence of the surface tension, which may be due to temperature or concentration gradients along the surface.

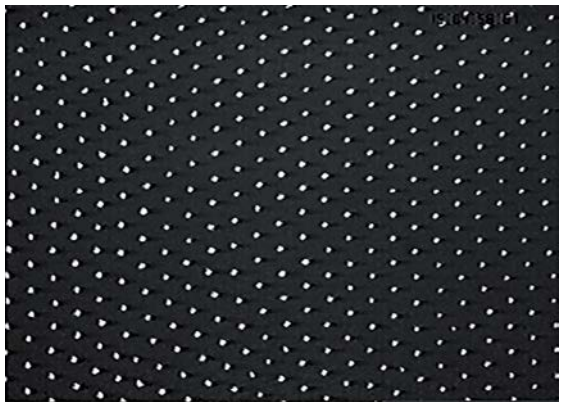


Figure 7.5.5: *An (almost) perfect plasma crystal viewed from above (Source: MPE Garching).*

the naked eye. Thus the melting and solidification process of a plasma crystal could be observed and the structure and frequency of lattice vibration could be measured.

The term **cold atoms** refers to atoms from which kinetic energy has been removed. They move correspondingly slowly. In a dilute gas of cold atoms they hardly collide. Their free path length has a macroscopic dimension and they move in a ballistic trajectory in response to gravity. If one injects cold atoms into a resonator, one can construct highly precise **atomic clocks** with them by using their resonance frequency as frequency norm. The residence time of the atoms in the resonator determines the quality of the time signal. Since atoms also react to gravity, their residence time in the resonator depends on the level of ambient gravity. Therefore, these clocks function best under microgravity conditions.

7.5.4 Solidification

The first-order phase transition from a liquid to a solid state is called solidification. It is of great technological significance since most raw materials are obtained from the liquid state. In the semiconductor industry and in optoelectronics it is essential to obtain from the melt, or sometimes from the gaseous state, single crystals which are as perfect as possible. In this context one speaks of **crystal growth**. Many metallic end and semifinished products are cast or molded; the solidification of metals is thus an important topic.

In solidification it is important that the melt in front of the solid–liquid phase boundary, the **solidification front**, be as homogeneous as possible, because any kind of inhomogeneity at the solidification front would leave its traces in the solid. This refers to both temperature and concentration gradients. If convection cells are present, be they stationary or time dependent, this homogeneity is no longer assured. Perfect single crystals can best be grown under purely diffusive conditions, that is in the absence of any flow. Microgravity offers optimal conditions for the purpose.

7.5.4.1 Crystal Growth

Crystals only grow if there is a driving force, namely a temperature and/or concentration gradient. Under microgravity conditions the Rayleigh number can be kept below its critical value so that these gradients do not lead to the formation of convection cells. However, free surfaces must also be avoided during convection-free crystal growth, since Marangoni convection would otherwise be inevitable. In the **floating-zone process**, which is common for semiconductor crystal growth, this is not the case. In this process a polycrystalline, cylindrical work piece is drawn through a narrow heating zone, melted and recrystallized. In the heating zone a free liquid surface is formed which leads to Marangoni convection. By coating the work piece with a thin layer, usually consisting of a high-melting oxide of the source material, the free surface and thus Marangoni convection can be avoided.

A special case of crystal growth under microgravity conditions is **protein crystal growth**. Large protein crystals are required for structural analysis by neutron scattering. The crystalline structure of protein crystals is quite complex; that is, the unit crystal cell from which the crystal built is relatively large. Under terrestrial conditions the protein crystal is deformed by its own weight while it grows and the regular structure essential for structural analysis is interrupted. Under microgravity conditions it was possible to grow protein crystals superior in size and perfection to those grown on Earth.

7.5.4.2 Directional Solidification

The **Bridgman process** is the most common process for so-called directional solidification in which the solidification front is forced in a particular direction by an imposed temperature gradient. Here the polycrystalline material is placed in a cartridge which moves relative to two heating elements which generate a linear temperature profile. The solidification front is where the temperature is just equal to the melting point of the material. The optimal arrangement is the vertical Bridgman process in which the temperature gradient is exactly opposed (antiparallel) to the force of gravity, where there is stable layering. If radial temperature gradients could be avoided, then this arrangement would be free of convection even under terrestrial conditions. But if one is dealing with alloys in which concentration gradients as well as temperature gradients play a role, a situation can arise in which there is no stable layering. In this case an analysis of buoyancy convection must consider the concentration dependence of the density. In addition to the thermal expansion $\partial\rho/\partial T$, the concentration dependence of the density $\partial\rho/\partial c$, plays a role. If both derivatives are of opposite sign, then stable layering is impossible and thermosolutal (**buoyancy**) convection takes place. Here the only remedy is microgravity.

7.5.4.3 Metals

Among metallic systems, the so-called **monotectics** deserve special attention. These are systems in which there are miscibility gaps in the liquid phase; that is, at particular temperatures and concentrations they separate into two immiscible liquid phases (Figure 7.5.6). A prominent example is Al–Pb. Such systems are, in principle, of considerable technical interest. If it were possible to finely disperse the two immiscible liquid phases and solidify them in that condition, then one could produce an ideal, self-lubricating material for **bearings**, whereby the hard phase takes over the support function and the soft phase the lubrication. This is of course not possible on Earth because the two separated phases segregate according to their differing densities; that is, the heavy lead sinks and the light aluminum floats to the top. It was hoped that under microgravity conditions this effect would not exist

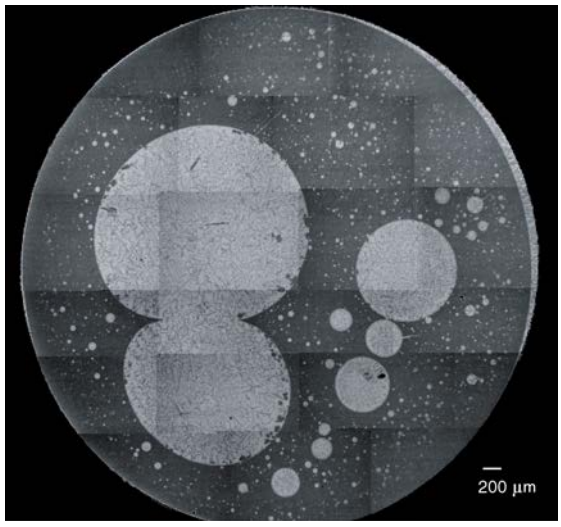


Figure 7.5.6: *Polished section of an (almost) completely separated CuCo alloy. The light regions are rich in copper, the dark ones rich in cobalt (Source: DLR).*

and a fine dispersion could be produced. But, as in the case of Marangoni convection, surface tension is again the spoilsport. A fine **dispersion** of liquid droplets of the one component in the liquid matrix of the other component has a large interface which contributes to the total energy of the system because the boundary surface energy between both liquids is retained. This energy is reduced by the **coagulation** of many small droplets into a larger one. Therefore, the equilibrium configuration consists of a large drop of liquid surrounded by the second liquid. Which of the two liquids is on the outside or inside depends on which of the two does a better job of wetting the container. One way to avoid this undesirable result is to quench the two-phase disassociated liquid melt before coagulation begins.

7.5.4.4 Undercooling

The cases discussed so far dealt with solidification near thermodynamic equilibrium. But since solidification is a first-order phase transition, the onset of the solid phase requires a critical nucleus. If it is possible to suppress or at least delay **nucleation**, then it is possible to cool the melt below its freezing point without solidification. The result is **undercooling**. This metastable condition of an undercooled melt and its solidification

is another fascinating field of materials science research under microgravity conditions. Here it is first necessary to achieve an undercooled condition and then to maintain it for a time. This means reducing the rate of nucleation. Besides the unavoidable homogeneous nucleation which is intrinsic to any system, there is also so-called heterogeneous nucleation, which is responsible for initiating solidification. Impurities and crystalline structures, as found for example on crucible walls, act as heterogeneous nuclei. Although the number of heterogeneous nuclei in a melt is much lower than the number of homogeneous nuclei, it is heterogeneous nucleation which prompts solidification. In order to study undercooling, undercooled melts and their solidification, **containerless processes** have been developed in which the melt can be positioned and processed without contact with a crucible. **Levitation melting**, in particular electromagnetic levitation, is one of these processes. Electrically conductive samples are caused to levitate by means of a high-frequency spatially inhomogeneous electromagnetic field and are then melted. The Lorentz forces act as a levitation force and the ohmic losses of the turbulent flow induced in the sample provide the heat. This process is routine under terrestrial conditions and allows undercooling of several hundred degrees over long intervals (minutes to hours). As a consequence of magnetohydrodynamic interactions the induced flows also lead to convection and sometimes to turbulent flow in the sample. Under microgravity conditions the positioning field can be reduced because it no longer has to levitate the sample but only stabilize it against the residual gravity. This in turn greatly reduces flow in the sample.

An undercooled melt is far from a condition of equilibrium. Its rate of solidification increases as the undercooling progresses. There are a number of theoretical predictions for the precise functional relationship between solidification rate and undercooling. Experimental confirmation requires a convection-free undercooled melt, something that in the ideal case can only be managed under microgravity conditions.

If the solidification of an undercooled melt has already begun, then the system is in a paradoxical situation: the solidified solid is warmer than the melt which surrounds it. To achieve equilibrium, the solid must give up heat to the melt. This is most successful if the interface to the melt, the solidification front, is increased.

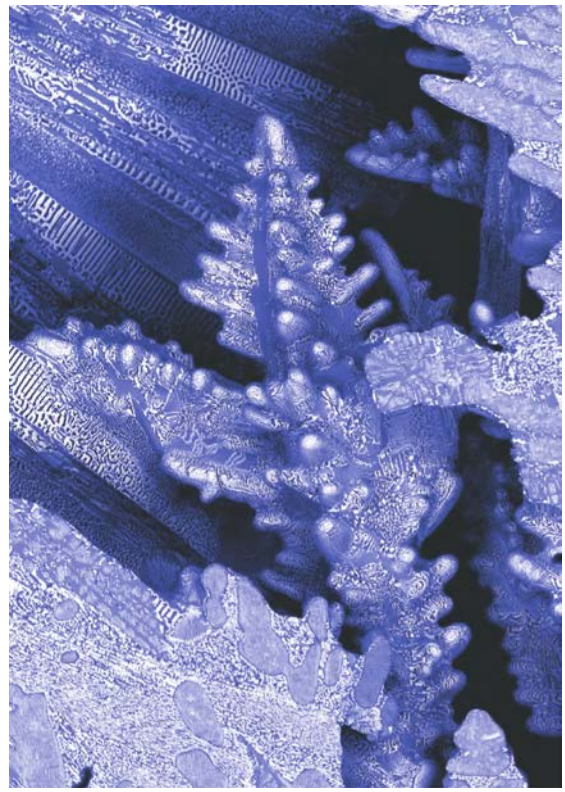


Figure 7.5.7: *Polished section of a dendritically solidified alloy (Source: DLR).*

Morphological instability is the result, namely a transition or mushy zone is formed, in which **dendrites**[6] arise (see Figure 7.5.7). Dendrites are fractal structures with a very large surface. Their form and growth can be theoretically calculated. For experimental confirmation it is important that convection either be suppressed or its influence be taken into account mathematically.

Another feature of solidification from undercooled melts is the fact that the phase diagram describing the existence range of the solid and liquid equilibrium phases is no longer valid. In the case of nonequilibrium solidification, metastable solid phases can arise if their growth rate is higher than the growth rate of the equivalent equilibrium phase. By means of **phase selection**, solids can be produced which have different characteristics than they would have if produced by

[6] A dendrite has a treelike structure, from the Greek δενδρον, a tree.

equilibrium solidification. An extreme case is glassy solidification, in which the liquid phase solidifies amorphously, that is without crystallizing.

7.5.5 Thermophysics

This section discusses how the ambient condition of reduced gravity can be used to precisely measure the thermophysical properties of liquids. In many cases these measurements are only possible under microgravity; in other cases measurements under microgravity improve accuracy.

Specifically, measurement of the following values is discussed:

- Density ρ
- Specific heat c_p
- Viscosity η
- Surface tension γ
- Electrical conductivity σ
- Heat conductivity λ
- Diffusion D
- Soret coefficient S.

Experiments under microgravity have been conducted on a large number of different materials under various conditions. These include aqueous solutions, melts of oxide glasses, high-temperature-fused salt baths, critical and supercritical fluids, and liquid semiconductors and metals. This section concentrates on liquid metals and semiconductors, primarily because of their technological relevance.

7.5.5.1 Measurements in Cartridges

Measuring Heat Conductivity

The heat conductivity of InSb was measured during the TEXUS-24 mission with the **transient hot wire method**. A short electrical impulse is sent through a thin wire located in the center of a cylinder (in this case filled with InSb). At the cylinder wall the temperature increase is measured as a function of time. If the heat is only transferred by conductivity, then the heat conductivity can be calculated from the signal as follows:

$$\lambda = \frac{Q}{2\pi} \cdot \frac{\mathrm{d}(\Delta T)}{\mathrm{d}(\ln t)} \qquad (7.5.11)$$

with Q the emitted amount of heat.

The advantage of the procedure is that the convective contribution to the heat transport is noticeable as a deviation from the linearity of the quotient $\mathrm{d}(\Delta T)/\mathrm{d}(\ln t)$. Under conditions of microgravity the entire range is linear, so the above quotient can be determined with high precision.

Measuring Self-diffusion and Interdiffusion

During the Spacelab missions (FSLP, D1 and D2) the self-diffusion and interdiffusion of several liquid metals like Sn, Pb, In and Sb were investigated. Also for these experiments it was essential that convection be avoided. The astonishing result was that the temperature dependence of the diffusion constants does not follow an exponential Arrhenius law, but can be described by a quadratic relationship:

$$D = A\,T^2 \qquad (7.5.12)$$

Such temperature dependence is theoretically predicted, but it conflicts with the widespread (prejudiced) opinion that diffusion is an activated process. In the context of the MSL-1 mission, a Japanese team repeated the experiments with tin. Their results confirm the power law behavior of diffusion, but with an exponent of 1.8.

Measuring the Soret Effect

The transport equations of linear irreversible thermodynamics predict that there are also nondiagonal **Onsager coefficients**, which link mass flows j_c with temperature gradients ∇T. This coefficient is called the Soret coefficient S and is defined by the equation

$$j_c = -S\,\nabla T \qquad (7.5.13)$$

For the cobalt-in-tin system, this coefficient was determined during the Spacelab missions FSLP and D1. In contrast to terrestrial measurements in which convection destroys the developing concentration gradient, the Soret effect is clearly perceptible and the Soret coefficient can be determined with high precision.

7.5.5.2 Containerless Experiments

Contact-free measurements of thermophysical properties under weightlessness became possible with

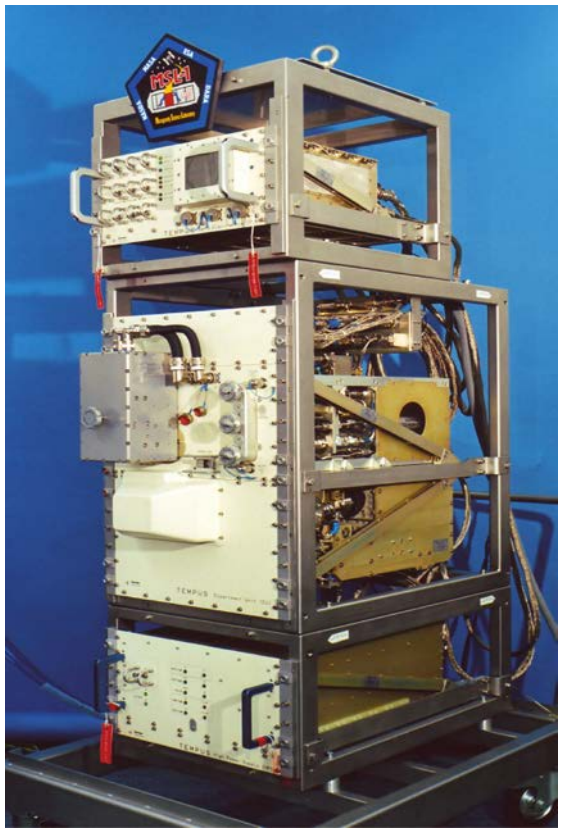


Figure 7.5.8: *Flight model of TEMPUS before integration (Source: EADS).*

the construction and use of the **TEMPUS** facility (Containerless Electromagnetic Processing Under Weightlessness facility). Because only small supporting forces have to be employed under weightlessness, the negative effects of the levitation field on precision measurements can be minimized. In particular, TEMPUS has the following advantages:

- Low heat input
  ⇒ No cooling gas required, so contamination is avoided.
- Minimal magnetic pressure
  ⇒ Spherical shape of the specimen.
- Low mixing effect
  ⇒ Laminar flow.

TEMPUS (Figure 7.5.8) was used on the two Spacelab missions IML-2 (1994) and MSL-1, MSL-1R (1997), during which an international research team measured a number of thermophysical parameters. These experiments are briefly described below.

Surface Tension and Viscosity

A free-floating drop of liquid exhibits **surface vibrations** around its spherical equilibrium form. The reaction force for these vibrations is provided by the surface tension γ. These vibrations are damped by the viscosity η of the liquid.

The relationship between the angular frequency ω and the surface tension γ is given by the following equation:

$$\omega^2 = \frac{32\,\pi\,\gamma}{3\,m} \tag{7.5.14}$$

whereas the damping Γ relates to the viscosity η as follows:

$$\Gamma = \frac{20\,\pi\,R_0\,\eta}{3\,m} \tag{7.5.15}$$

In the above equations m is the mass of the droplet. Their application is limited to space experiments where the requirement of a force-free droplet is easily fulfilled. The strong levitation field required in terrestrial experiments leads to a deformation of the droplet and splitting and shift of the stated frequency. The surface tension of a congruently melting $Au_{56}Cu_{44}$ alloy was measured during the IML-2 mission under microgravity conditions.

Viscosity cannot be measured contact-free on Earth, because the levitation field generates turbulent flows in the specimen which hide the damping of surface vibrations by the viscosity. However, it was possible to measure the viscosity from the damping of surface vibrations during the MSL-1 mission. The experiments were carried out using the eutectic alloy $Pd_{76}Cu_6Si_{18}$.

Electrical Conductivity

The measurement principle is based on the fact that the levitated droplet is part of the oscillating circuit which supplies power for the levitation coil. A change in the electrical conductivity of the sample accordingly leads to an **impedance change** of the oscillating circuit, which is identified by a change in the effective

resistance U_0/I_0 and by a phase shift φ. These values can be measured. The relationship to conductivity is given by the following equation:

$$\sqrt{\frac{2}{\omega\,\sigma\,\mu_0}} = \frac{R_0}{2}\left[1 - \sqrt{1 - A\left(\frac{I_0}{U_0}\cos\phi - B\right)}\right] \quad (7.5.16)$$

with ω the frequency of the oscillating circuit, μ_0 the magnetic permeability and R_0 the radius of the sample; A and B are constants. This simple relationship is only true for spherical samples in a homogeneous magnetic field, which limits the application of this method to experiments under microgravity conditions. During the MSL-1 mission the electrical conductivity of a $Co_{80}Pd_{20}$ alloy was determined in this manner.

Density and Thermal Expansion

If the mass is known, then the density of a sample can be determined by calculating its volume. Under microgravity conditions liquid drops are spherical, so the volume calculation is reduced to the measurement of the diameter of the specimen. This was successfully accomplished during the MSL-1 mission for the glass-forming zirconium-based alloy ZrCuNiAlCo.

Specific Heat

Using **modulation calorimetry** the specific heat c_p of levitated samples can be measured. The precondition is that the heat balance is precisely known, which means that the heat input and output must be exactly measured. The heat input depends on the efficiency of the inductive heating of the sample and can be determined by measuring the electrical conductivity. The heat losses of the sample are known if the experiment takes place under UHV conditions so that only radiation losses occur. Under these conditions, there is the following relationship between the modulation amplitude of the heat output ΔP and the temperature variation ΔT:

$$\Delta P/\Delta T = \omega\, c_p, \quad (7.5.17)$$

where ω is the modulation frequency. As mentioned previously, the sample has to be convectively cooled with gas in the case of terrestrial experiments so this approach can only be taken under microgravity conditions.

7.5.6 Payloads

The experiments described above require experimental equipment which represents the actual payload, which is the reason for constructing and operating designated spacecraft.

Such equipment is usually a **multipurpose facility** because of its high production costs, which means that it is used for a variety of experiments and by different scientists.

These multipurpose facilities are usually built by the space industry. Construction by scientific institutions is the exception.

These payloads are customized for their particular platform or mission. For materials science experiments there are payloads for parabolic flight missions, high-altitude research rocket missions, especially in the German TEXUS program,[7] and human and unmanned orbital systems, including the ISS. The customers who order these payloads are the large space agencies NASA, ESA, DLR and JAXA. In the area of utilizing microgravity, DLR is especially active, whereas NASA discontinued the construction of such payloads years ago.

An experimental facility basically consists of the following components:

- Processing chambers
- Actuators
- Sensors
- Peripheral equipment
  - Power supply
  - Electronics
  - Data management.

In the processing chamber the experiments run under defined ambient conditions, that is temperature and atmosphere (vacuum, inert gas, etc.) are prescribed. Various activation mechanisms initiate the process under investigation, which is diagnosed by various sensors. These include, for example, temperature probes and optical systems such as video cameras. The

[7] TEXUS: Technology Experiments Under Weightlessness.

experiment is controlled by astronauts, interactively from the ground (**telescience**) or automatically. The data obtained is either stored on-board or transmitted to Earth in real time or after a delay.

Every payload has various interfaces to the platform relating to:

- Structure
- Power supply
- Cooling
- Gas/vacuum
- Analog data
- Digital data.

Materials science payloads are usually furnaces or systems to manipulate fluids at room temperature. Furnaces place high demands on the power supply and cooling equipment, whereas the investigation of fluid systems usually generates a high throughput of digital video data.

Some of the currently existing or specifically planned materials science payloads for the ISS are discussed below.

Microgravity Science Glovebox (MSG)

The MSG (Figure 7.5.9) was transported to the ISS in 2002 and has been in operation ever since. The glovebox has a working volume of 255 liters and is equipped with an airlock. It can be flooded with inert gas (nitrogen) or evacuated and is provided with four video cameras. The MSG provides optimal working conditions for small, autonomous payloads whose operation requires a controlled environment.

European Drawer Rack (EDR)

The EDR provides the infrastructure for small payloads which do not require their own rack. They are accommodated in drawers or lockers. The EDR can supply $5 \times 280\,W + 1 \times 600\,W$ of electrical power for the individual payloads and provides digital and video data transmission. It also provides cooling power up to 2200 W. At first, accommodation of the FASTER (Facility for Absorption and Surface Tension on European Rack) and PCDF (Protein Crystallization Diagnostic Facility) experiments is planned, and later the MSL-EML (Materials Science Laboratory – Electromagnetic Levitator).



Figure 7.5.9: *Microgravity science glovebox (Source: ESA).*

Materials Science Laboratory (MSL)

The MSL is the workhorse of materials science experiments on the ISS. It consists of a double rack and will be accommodated in the Columbus module as part of the initial equipment of that module. Because of its size it must be transported on the shuttle. The MSL provides the required infrastructure (power, cooling, data transfer, etc.) for various furnaces which are designed as plug-in units for the MSL (Figure 7.5.10). At the present time there are two furnace modules, namely:

- LGF (Low-Gradient Furnace)
- SQF (Solidification and Quenching Furnace).

The samples for both furnaces are housed in cartridges which also contain the required diagnostic equipment (thermal elements etc.). At the same time they serve as containers for transporting the samples to and from the ISS. They weigh several kilograms and have to be correspondingly taken into account during mission planning.



Figure 7.5.10: *Engineering module of the Materials Science Laboratory (Source: ESA).*

The LGF is a furnace for melt and solidification experiments that require either isothermal conditions or merely low temperature gradients. It consists of two heating zones separated by an adiabatic zone.

The SQF is a gradient furnace for directional solidification. It too is resistor heated but contains a water-cooled zone and a ring of liquid metal which makes possible temperature gradients to about 150 K/cm.

In addition, the MSL has a quenching device which can transport the samples from the hot to the cold zone at a maximum 100 mm/s in order to freeze the structures which arise during solidification and avoid any subsequent transformation of the solids. The maximum temperature for both furnaces is 1673 K.

Programmatically belonging to the MSL but physically a separate entity is the MSL-EML. The EML is an electromagnetic levitation device to process metallic melts without using a container. It was developed from the German TEMPUS facility, successfully employed in three Spacelab missions. Originally designed as a rack assembly for Columbus, the EML was later broken down into modules which can be integrated into the European Drawer Rack (EDR). This makes it possible to transport the facility with the ATV to the ISS. This measure became necessary because, at the time of its launch (2011), the shuttle will not be available as a transport system.

The EML has a coil system optimized for utilization under microgravity conditions which on the one hand decouples positioning and heating and on the other hand has highly efficient inductive heating and positioning power. The maximum power of the EML is 2500 W, with which temperatures in the sample of about 2300 K can be achieved. The samples are in containers which prevent possible contact of the liquid sample with the levitation coils. Seventeen samples and their containers are accommodated in a sample carousel. The carousel also serves at the same time as a transport container for the samples to and from the ISS and weighs about 17 kg, which is equivalent to a mass of about 1 kg for each sample to be transported. The transport of the sample is accomplished either under vacuum or in a protective atmosphere.

In addition to its high power consumption the EML generates also high data rates, in particular video data. These are needed on the one hand for process



Figure 7.5.11: *The Fluid Science Laboratory (Source: ESA).*

control in real time on the ground, and of course they also provide the necessary data for scientific evaluation.

Fluid Science Laboratory (FSL)
The FSL (Figure 7.5.11) makes it possible to conduct experiments at room temperature with liquids. The liquid volume can be up to 1000 cm^3 and is located in a drawer unit typically weighing 30 kg. The FSL has numerous optical diagnostic systems, such as high-resolution, high-speed video cameras in two directions, Schlieren photography, interferometry and holography.

Bibliography

[7.5.1] Feuerbacher, B., Hamacher, H., Naumann, R. (eds.) *Materials Sciences in Space*. Heidelberg: Springer Verlag, 1986.

[7.5.2] Walter, H. (ed.) *Fluid Sciences and Materials Sciences in Space*. Heidelberg: Springer Verlag, 1987.

[7.5.3] Ratke, L., Walter, H., Feuerbacher, B. (eds.) *Materials and Fluids under Low Gravity*. Heidelberg: Springer Verlag, 1996.

[7.5.4] Keller, M., Sahm, P. (eds.) *Bilanzsymposium Forschung unter Weltraumbedingungen*. Aachen: WPF/RWTH, 2000.

[7.5.5] Seibert, G. (ed.) *A World without Gravity*, ESA SP-1251. Noordwijk: ESA, 2001.

[7.5.6] Lämmerzahl, C., Dittus, H. Fundamental Physics on the ISS. *Gen. Relativ. Gravitation*, **36** (3), Special Issue, 2004.

[7.5.7] Feuerbacher, B., Stoewer, H. (eds.) *Utilization of Space*. Heidelberg: Springer Verlag, 2006.

[7.5.8] Bárczy, P., Roósz, A. (eds.) *Solidification and Gravity I–IV*. Zurich: Trans Tech Publications, 1991–2006.

7.6 Space Medicine and Biology

Hanns-Christian Gunga, Mathias Steinach, Karl August Kirsch and Andreas Werner

From operational, physiological and medical points of view, long-term missions in space can be characterized by

- Flight duration
- High mission autonomy
- Varying gravitational loads (hypergravity during launch and landing, weightlessness during interplanetary transit, hypogravity during presence on the Moon, artificial gravity loads)
- Psychological, physiological and social problems
- Cosmic radiation
- Hypomagnetic environment.

These exceptional natural and artificial environmental conditions during a long flight in space in a space vehicle or during a stay at a lunar or Martian base can cause physiological, anatomical, biomechanical, psychological and social changes, and have a lingering influence on the organism in complex ways. These include performance, health and the general sense of well-being of the astronaut (sense of taste, tactile sense, reaction time, coarse and fine motor functions, muscle strength, etc.) as well as crew safety, health management and the training program. It must also be recognized that seemingly marginal medical, physiological, psychological or social changes which occur during short flights can develop into pathological problems during long space flights and thereby threaten the entire mission.

7.6.1 Medicine in Space

In the second edition of this handbook the problems associated with human ventures into space were extensively and thoroughly presented based on the state of knowledge at the time, with special emphasis on the technological problems. That presentation can still serve as a useful reference, so the present text will

concentrate on the latest developments. Then, the ISS was anticipated as the endpoint for the time being in space flight planning. Seen with today's eyes, those plans need to be supplemented with a look at the problems associated with **permanent colonization of the Moon** and with a **human mission to Mars.** This means to consider stay-times in space which are measured in years, which gives rise to significant and in some cases entirely new demands. Disregarding the very high financial expense, and not mentioning the technical difficulties, very thorough cost–benefit assessments would be required to convince the public of the relevance of such an endeavor. So far, the cost–benefit analyses have made sense in terms of value received. In this context just the one example of the scientific yield for astronomy of the **Hubble Space Telescope** should be mentioned, which helped to revolutionize basic astronomical conceptions. The Hubble Space Telescope was launched in 1990 but after three years had to be modified by astronauts during extravehicular activities to eliminate an optical defect. In 1998 it was again serviced by astronauts. In 2008 it is actually due to be decommissioned, unless another service mission is undertaken. In the meantime NASA has approved such a mission. This example may illustrate how important a human presence is in space. In this chapter the emphasis will be on the problems connected with **very long human missions,** extending beyond the ISS.

As can be seen in Figure 7.6.1, there is a rather good database for approximately the first 30 days of habitation in space.

In the meantime we know from numerous investigations how the physiological progression of various processes takes place during this phase, for example in the **cardiovascular system** and the thermoregulatory system and concerning the **salt/water balance.** Afterward, a stable equilibrium is reached after about six weeks (Figure 7.6.2).

However, because of the scarce amount of relevant data we know relatively little about the **muscle and skeletal system** during long periods in space, and this despite the fact that for very long stays of humans in space it will likely play a significant role. The same is true for long-term effects of increased **radiation exposure** in space. The progression and extent of recovery from this radiation load and the changes to

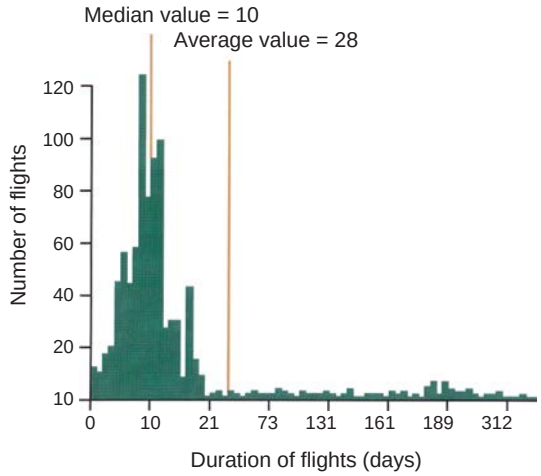


Figure 7.6.1: *Frequency and duration of human space flight. Note the almost complete lack of long-term flights. Median flight length is only 10 days (Source: [7.6.3]).*

other human organ systems after **reentry in Earth's gravitation field** are also open questions. And it can be expected that during long flights in space, psychophysiological problems and interpersonal conflicts will be an increasing problem. A number of **stress factors** affect humans in space, including on the one hand isolation, confinement, lack of sources of stimuli, dependence on technical systems and lack of comfort, and on the other hand individual (lack of privacy, separation from family members, stress management, motivation, etc.) and interpersonal (territorial and dominance behavior, group structures, group thinking, group dynamics) conflicts, but also the leveling of circadian, lunar, circa lunar and annular endogenous rhythms. During long flights in space these changes are combined with an increasing feeling of monotony the longer the flight lasts. All these factors contribute to making a long-duration flight in space, even after years of training, an unusual burden for astronauts that is far beyond their range of experience on Earth.

7.6.2 Mission Scenarios

Human exploration of space requires that the health and performance of astronauts be assured in different mission scenarios:

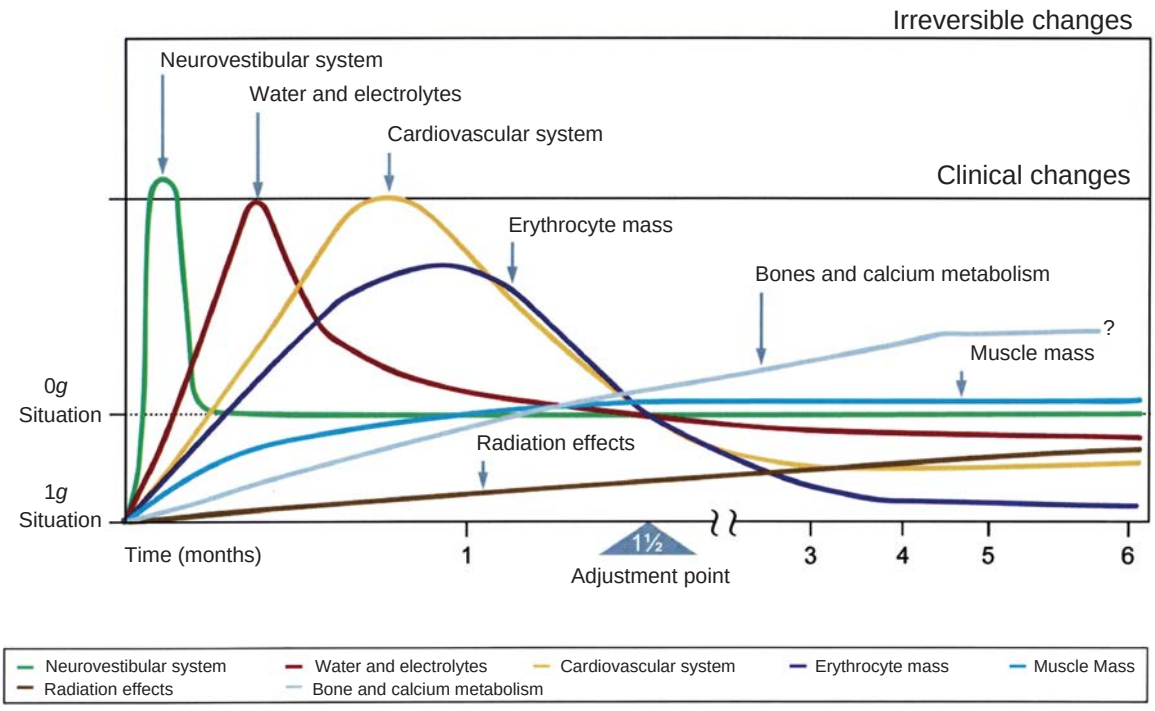


Figure 7.6.2: *Temporal progression of physiological changes during adjustment to weightlessness. The 1g condition represents the physiological situation on Earth. The 0g (zero-gravity) condition is full physiological adjustment to space conditions, which probably could only be achieved by those born in space. An astronaut requires an average of six weeks to reach a point of adjustment at which there is at least partial adaptation to the new environment (Source: [7.6.15]).*

1. In near-Earth orbits (ISS, low Earth orbit – LEO)
2. During an interplanetary flight (transit)
3. On the Moon
4. On Mars.

On the basis of these scenarios, the following parameters can be identified which fundamentally influence the mission and which must be taken into account to varying extents depending on the nature of the particular mission:

(a) **Distance** from the Earth
(b) Changes in **g-forces** (10^{-6}g in LEO, 0g during an interplanetary transit, 0.16g on the Moon, 0.38g on Mars), as well as
(c) The probable **duration** of a mission.

These mission scenarios lead to various load profiles related to health, safety and the well-being of the crew. In addition, they significantly influence operational planning and technological issues down to the construction and design of the requisite space vehicles or planetary space stations (habitat) [7.6.1], [7.6.2]. In general, autonomy as well as safety and health problems substantially increase from scenario 1 to scenario 4 because:

1. A **real-time telecommunications system** becomes more and more inefficient with increasing distance from the Earth. For scenario 4 (Mars) there is a time delay of 40 min, so that a natural dialog between the crew on Mars and the ground station on Earth can hardly take place.
2. An **evacuation** of astronauts from LEO in an emergency would be possible within 24 hours (Soyuz capsule), and could be managed from the Moon in the course of several days/weeks, but timely evacuation is as good as impossible in

scenarios 3 (transit) and 4 (Mars), which means that the highest degree of autonomy is required for these scenarios.
3. Finally, the **radiation risk** in scenarios 2 to 4 substantially increases without appropriate shielding measures.

7.6.3 Experience to Date

So far, about **460 people have been in space,** most of them male astronauts; the female proportion is only about 10%. The cumulative flight time is about 80 years (as of August 2007). The frequency and length of human flights in space (Figure 7.6.1) point out the noteworthy fact that the median flight duration is only 10 days. Until 2003 only about 40 astronauts had flight experience longer than six months, and only four people were in space longer than one year [7.6.3]. The record for the longest continuous presence in space is still held by the Russian physician Dr. Valeri Polyakov who flew for **437 days** on the MIR station from 1994 to 1995. The cumulative records are also held by Russian cosmonauts, namely Valeri Polyakov (679 days, two flights) and Sergei Avdeyev (748 days, three flights). Experience in space is always supplemented with **accompanying terrestrial studies,** something which often passes unnoted. These studies have always been a major component of space medicine and it can be safely assumed that their significance will increase for the planning and execution of future long flights into space.

It should be emphasized here that the European Space Agency (ESA) critically analyzed at an early date the various **medical, physiological, psychological, social, and technical problems** which could play a role during lengthy human presence in space. This subject has been exhaustively treated in ESA studies, at special meetings and in numerous technical reports [7.6.4], [7.6.5], [7.6.6], [7.6.7]. In this context one also speaks of the **human factor** which must be considered in human missions. By definition, studies of the human factor are an interdisciplinary research field which integrates insights from medicine, physiology, psychology, sociology, biomechanics, engineering and other disciplines, and complements them with **habitability studies.** These studies concern human abilities and maximum potentials. They undergird the design of equipment (ergonomics), machines, systems, workflows, tasks, and ambience so that people find them as safe, comfortable and effective as possible. Because the necessary information cannot be adequately appreciated and dealt with only on the basis of theoretical considerations, ESA, for example, established the interdisciplinary Simulation Mission Study Group (SIMIS) several years ago. It was the task of this working group to identify **simulation facilities** (pressurized diving chambers, submarine stations, bunkers, etc.) or other isolated places on Earth where people in small groups can live together to simulate the living quarters of long-duration flights in space. It turned out that polar stations, underwater research stations, hyperbaric and hypobaric chambers, bed-rest facilities, submarines, immersion facilities, parabolic flights and mock-ups (habitability test bed) all can serve as reasonably analogous environments. In the meantime ESA has conducted numerous **isolation and confinement studies,** and results are available from simulation studies making use of various facilities with different crew compositions and crew sizes (ESA reports: POLAREMSI, ANTEMSI, HYDREMSI, ISEMSI, EXEMSI, HUBES, SFINCSS). They make far-reaching recommendations on planning and conducting long-duration missions from medical, psychological and technical perspectives [7.6.4], [7.6.5], [7.6.6], [7.6.7]. These terrestrial simulation studies will continue to be a significant element of all future space missions.

It is astonishing that there is hardly any data available from the field of clinical medicine, specifically on problems related to isolation and confinement, although similar problems must arise in cases of long-term immobility (fractures), for example. A current example for the clinical relevance of results from such simulation studies are the findings of Heer [7.6.8] and Titze [7.6.9], who could prove that the human body can store osmotically inactive sodium. This offered a completely new perspective on the endemic disease of high blood pressure (hypertension). It was significant that the discrete retention of salt was only noted after observation periods of three to eight weeks. Short-term observation periods of three to five days were inadequate to reveal this effect. In the meantime it was experimentally shown that the surface tissues (cutis

and subcutis) play a central role in the osmotically inactive storage of sodium.

Also, as mentioned above, problems of isolation and confinement have been the subject of intensive research in space medicine for years and are leading to results that could easily be relevant for clinical problems, considering that almost half of the population in industrialized countries lives in single-person households, either by necessity (sickness) or voluntarily (singles). Here, too, medical, physiological space research on humans is setting the pace in describing the physiological phenomena and pathophysiological changes which arise. Furthermore, a **methodological repertoire** from space research is made available which can later find practical application in clinical medicine, such as noninvasive technologies, telemetric transmission and real-time data analysis programs.

7.6.4 Environmental Parameters

Table 7.6.1 lists a number of **life support systems in space** (LSS) which have been developed on the ground and only in some cases tested under weightless conditions. All these systems have a protective component in that they, for example, remove CO_2 or produce O_2 [7.6.10]. At the same time they have a life support component since they are supposed to provide energy-rich substrates for consumption. As can be seen from the designations in Table 7.6.1, they vary in nature according to the purpose they are supposed to fulfill in space.

Bioregenerative life support systems (BLSSs), if they are to fulfill their purpose, depend on the photosynthetic activity of green plants, which in the presence of light convert CO_2 and water into carbohydrates and oxygen, providing chemical energy. In order to be fully effective, in addition to sufficient concentrations of CO_2 and O_2 in the ambient air and the right humidity, temperature and barometric pressure, light of a suitable wavelength must also be available, all of which poses difficulties in space. The circadian and seasonal rhythms of plants must also be taken into account. Cultivation in these artificial systems is usually in the form of monocultures, so the plants lack a natural social milieu (biotope), which, as is sufficiently well known from terrestrial experience, involves risks.

The lack of gravity is the biggest disadvantage, because all systems developed to date use gravitropic land plants. These do not prosper under weightlessness, if they grow at all, and do not produce edible fruits. The cause is likely the lack of thermal convection around plants growing under weightlessness. For this reason, the CEBAS system, among others, was developed, which produces biomass in an aquatic milieu using nongravitropic plants and can provide animal protein as well as plant material [7.6.10]. Despite all these problems, such systems have to be further developed until satisfactory solutions are found. Self-contained, **completely closed systems** which can be harvested will probably not exist for the time being; in most cases something has to be either added to the system somewhere or removed from somewhere else. The systems are **semi-open,** which means it costs energy to maintain them. One can, however, develop energy-saving concepts for them. From the perspective of physiologists and medical doctors it should also be noted that for long-term presence in space a good variety of energy producers and protein-rich nourishment to guard against a certain monotony in diet must be assured. Otherwise, dietary deficiencies and malnutrition can be expected in the long term.

Table 7.6.1: *Life support systems in space.*

CEBAS	Closed Equiliberated Biological Aquatic System
BLSS	Bioregenerative Life Support System
ARS	Atmosphere Revitalization Subsystems
ECLSS	Environmental Control and Life Support System (Space Shuttle)
ECLS	Environmental Control System
CELSS	Controlled Ecological Life Support Systems
CEEF	Closed Ecology Experiment Facility

7.6.5 Medical Physiological Problems Arising from Residence in Space

Various authors [7.6.3], [7.6.11], [7.6.12] have summarized the most relevant medical physiological experience, problems and risks that have become known in the course of US and Russian short- and long-term

flights in space, as well as in work with terrestrial simulation models produced over the last 30 years. The conclusion is that when planning, designing and realizing missions in space the following **medical problems** must be taken into account: loss of appetite (anorexia), nausea (space motion sickness), exhaustion, sleeplessness, dehydration, dry and inflamed skin, muscular–skeletal pain in the back, legs and feet, respiratory tract infections, cardiovascular problems including arrhythmia, headaches, diarrhea, constipation, barotraumas (the formation of free gas bubbles in blood and tissues causing embolisms, limb and joint pains, called "bends"), chemically induced lung infections as a consequence of long, strenuous extravehicular activity (EVA), deficient sensory perception, seborrhea, allergic reactions, fungal infections, dental disease, down to the inhalation of foreign bodies.

In the compilations made by Billica *et al.* [7.6.11] and Clément [7.6.3], considerable attention is given to **psychological problems** which reduce the mental performance, ability to concentrate, psychic stability and resilience of astronauts (asthenia, stress, fear, nervousness, depression, etc.). The entire spectrum of possible problems is probably still not yet covered. Comet [7.6.7] has identified the following **research areas** which in the future will definitely require more work before a human flight to Mars lasting 500 to 1000 days can be considered. These include the performance of the cardiovascular, immune, muscle and skeletal systems, nutrition under complete weightlessness (0–0.1g), changing gravity loads (1/10g, 3/10g, 5/10g, 8/10g), toxicological and pharmacological kinetic studies under weightless conditions, studies on the occurrence of kidney stones caused by accelerated bone reduction in space, the dynamics of healing processes, as well as research on the influence of various types of artificial gravity on the human, animal and plant organism.

7.6.5.1 Changes in Body Composition

Liquids and electrolytes, fat-free mass (muscle mass), fat, and bone mass all undergo considerable changes during the first six weeks under simulated (terrestrial isolation studies) and actual weightless conditions [7.6.13], [7.6.14]. Biochemical equipment and technologies must be available on-board to identify these bodily changes. If changes occur, then this data is important for planning and initiating special treatment, for example as countermeasures in the case of progressive loss of bone mass during a space flight, such as intensified exercise on a vibration platform or training exercises with a flywheel ergometer.

7.6.5.2 Cardiovascular System

It has been confirmed in several studies that the cardiovascular system undergoes rapid changes during presence in LEO, on short-term (Space Shuttle) and long-term (Skylab, MIR and ISS) human missions [7.6.15], [7.6.16]. There is **fluid displacement** toward the head [7.6.17], [7.6.18], [7.6.19], which leads to a loss of liquid in the lower extremities of about 2 l and reduces the astronauts' **plasma volume** by up to 20% [7.6.15], [7.6.20], [7.6.21]. The production of **red blood cells** also decreases significantly during the first weeks in space, since erythropoietin production and release from the kidneys is reduced [7.6.22], [7.6.23]. The reduced plasma volume and lower red blood cell mass – the latter is critical for oxygen transport capacity – reduce the maximal exercise astronauts can undertake, especially after landing back on Earth. In addition, reduced plasma volume as well as an altered **sympathetic nerve receptor density** and threshold in the blood vessels and the heart can play a key role in orthostatic hypersensitivity. Also, the shift of volume along the body axis from the body's lower to upper half leads to noticeable changes in **heat emission via the skin** and probably affects the entire human heat balance under weightlessness. This is supported by recent data from relevant human physiology experiments in connection with parabolic flights (Figure 7.6.3).

Under terrestrial conditions the head and neck are responsible for over 30% of the heat exchange between the body and the environment. Forced convection to ventilate the space vehicle or the ISS could cause heat emission from this part of the body to be much higher. This has been empirically confirmed by the experience of numerous astronauts who reported cool to cold extremities, to the extent that some of them wore socks on their feet although the room temperature was relatively high (ISS, approx. 29 °C).

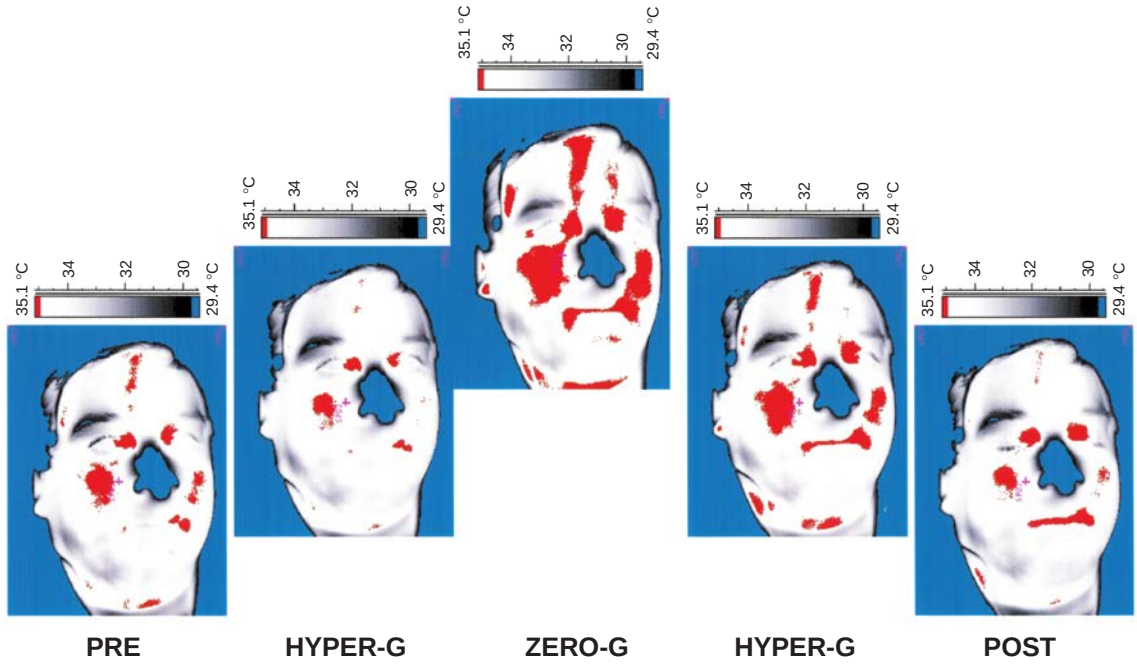


Figure 7.6.3: *Infrared image of facial skin temperature (male test subject, 20 years old) in the various phases of a parabolic flight during DLR's eighth parabolic flight program in 2006. The rapid dynamics of skin temperature changes which are caused by a fluid shift along the body axis and during weightlessness lead to a striking increase in heat emission (Source: DLR).*

7.6.5.3 Muscles and Skeletal System

Both the muscles and skeleton require enough stimulus in the form of mechanical loading ("strain") to be able to maintain their structure. Under weightless conditions this strain is reduced to a minimum. **Muscle atrophy** is noticeable after only a brief time in space, accompanied by structural and functional changes in the muscles as well as a negative nitrogen balance. This atrophy is especially pronounced in the posture muscles which support the body against gravity. Without corrective measures astronauts lose up to 20% of their muscle mass during a short mission and up to 50% during long-term missions [7.6.3]. This loss is higher in the group of extensor muscles compared to the flexors. The leg muscles have the highest loss since the astronauts move through their new environment mainly by using their arms. Together with the changes in the neurovestibular and cardiovascular systems, like dizziness and orthostasis, the loss of strength and stamina lead to a considerable reduction in performance for astronauts returning to an environment with gravity – which in the case of a landing on Mars could hinder the success of the mission.

Changes in muscles are closely linked to changes in the skeletal system. Without adequate strain there is an imbalance between bone development and breakdown with a global **loss of bone mass** of 1 to 3% per month on average. In areas experiencing especially low strain the bone density is reduced by up to 10% [7.6.3]. Loss of bone mass causes hypercalciuria. Although other factors besides increased calcium levels in the urine are relevant (pH value, uric acid, oxalate), there is an increased risk of **post renal stone formation** [7.6.3], [7.6.15]. Bone decomposition and the associated increased risk of fractures, particularly of the lower extremities, and possible renal stone formation seriously threaten the health of astronauts and the success of a mission. Appropriate countermeasures must be taken.

7.6.5.4 Sensory Systems

The integration of visual, vestibular and kinesthetic stimuli makes orientation through the environment possible. The individual systems are interconnected. Thus proprioceptive and visual information lead to the vestibular system and from there back again to involuntary control of eye movement. The lack of gravitational stimulation leads to an incongruence of information of the semicircular canal of the inner ear, which independent of gravity only reacts to a change of position and the acceleration information from the saccule and utricle. The discrepancy between this information and the kinesthetic and visual signals can lead to so-called **space motion sickness,** which is expressed in various types of dizziness, feeling ill or nausea [7.6.15], [7.6.24]. After a period of adjustment of a few days, these symptoms disappear, but can return after the astronaut is back in a 1g environment.

Some astronauts reported problems with **nearsightedness,** which may be caused by fluid displacement toward the head and changes in the shape of the eyeball and lens in a 0g environment [7.6.3].

Because of the various permanent sources of noise on a space vehicle (motors, experiments, air-conditioning equipment) **noise pollution** reaches levels of 60 to 100 dB [7.6.3], which besides causing stress also carries the risk of partial deafness and makes appropriate acoustic and auditory protection necessary.

7.6.5.5 Diet

Numerous factors must be considered regarding diet and nutrition. The food must be nonperishable (for months to years), the drinks powdered, the meals freeze-dried. Water is added and mixed into the food only on-board. Many meals are heated to high temperatures to make them long-lasting. All efforts must be made to avoid **food poisoning,** which could be life threatening for the entire crew on an interplanetary flight. In recent years it has been shown particularly during long-term residence on MIR and the ISS that astronauts love "special event meals" to counteract the **monotony** caused by external circumstances [7.6.3]. Therefore, attention must be paid to taste and even to the type of packaging. Of course, nutritional value and trace element content must be precisely known and adequate. Food that is too salty is to be avoided since it fosters an increased removal of calcium from the bones, which is accelerated anyway because of weightlessness.

NASA has a collection of about 150 menus, to which another 150 menus from the Russians can be added on the ISS, so the astronauts have about 300 different menu variations available. Alcohol is not permitted. The day's menu is divided into four segments: breakfast, lunch, dinner and a snack. However, the consumption of cookies is problematic since crumbs are not easy to remove under weightlessness and can have unpleasant consequences like crumbs in the eyes, in other body openings, interference with technical and scientific equipment. Carbonated drinks must also be avoided since they lead to unpleasant flatulence or eructation (the Latin word for belching). The latter can also sometimes lead to spontaneous vomiting under weightlessness.

Astronauts and cosmonauts have a marked preference for strongly spiced food. In part, the reason could be the liquid displacement along the body axis caused by weightlessness, which causes the oral mucosa in the mouth and the tongue to swell (edema). Also, the distribution of food in the mouth changes because of weightlessness, which leads to an altered **sense of smell.** All these factors taken together are probably sufficient to show that nutrition under conditions of weightlessness – particularly for long flights – is a problem which should not be underestimated. It has already been well confirmed in isolation studies that a balanced, varied diet is a decisive factor for the overall psycho-physiological well-being of astronauts, whereby communal meals involving the entire crew also play an important role in the daily schedule (the social components of meals), independent of energy supply and taste [7.6.3], [7.6.15], [7.6.25].

As far as energy requirements are concerned, it was surprising to discover that the **caloric demand** under weightlessness hardly differs from that in 1g conditions for the astronauts. This could be because there are increased energy requirements under weightlessness due to the daily muscle exercises instituted as a countermeasure to muscle and bone loss in space, as well as due to the forced convection by the air-conditioning system inside the spacecraft

because natural thermal convection is impossible under microgravity conditions. Changes in digestion (resorption) and taste could also play a role [7.6.26], [7.6.27].

7.6.5.6 Radiation

Clinical effects of exposure to radiation can be divided into short- and long-term effects. The basis for both is the formation of highly reactive free radicals by the radiation, which can damage cell components, particularly DNA and other control molecules. The **short-term effects,** also known as **radiation sickness,** involve symptoms like nausea, vomiting, diarrhea, reduced blood count, hemorrhaging and death, depending on the strength of the dosage. **Long-term effects** affect the DNA; radiation damage can lead to uncontrolled cell growth and finally to cancer if compensatory mechanisms like apoptosis (programmed cell death) or DNA repair (p-53 mechanism with premitotic pause and restoration) are ineffective [7.6.3].

The **radiation risk** to astronauts is significant. For missions near the Earth (LEO) crew members are well protected from ionizing radiation by the Earth's magnetic field. But further out, the crew is exposed to far higher radiation levels. In the Van Allen belt, which extends up to an altitude of 30 000 km, radiation composed of high-energy protons and electrons (approx. 1 keV to several megaelectronvolts for electrons and hundreds of megaelectronvolts for protons) prevails. **Cosmic radiation** (also known as cosmic background radiation) consists of high-energy ionized nuclei, from hydrogen to uranium nuclei (up to 1000 MeV), whose origin is outside our Solar System. **Solar particles** are another important source of radiation risk, primarily consisting of high-energy protons which are ejected into space during a solar eruption (solar particle event (SPE)). This radiation can significantly increase the energy of cosmic particles, up to several hundred megaelectronvolts per nucleus. Neither the temporal occurrence and duration nor strength of such an eruption has been reliably predicted so far, whereby the radiation from SPE represents a serious danger for interplanetary long-term missions [7.6.1], [7.6.3].

Radiation **dosage** D is measured in grays (Gy), where 1 Gy is equivalent to an absorbed energy of 1 J/kg. Since the various radiation sources also differ in the strength of the biological effects, the so-called equivalent dosage D_e in sieverts (Sv) is used. Depending on the radiation source, the radiation dosage D is multiplied by a factor for its significance in each

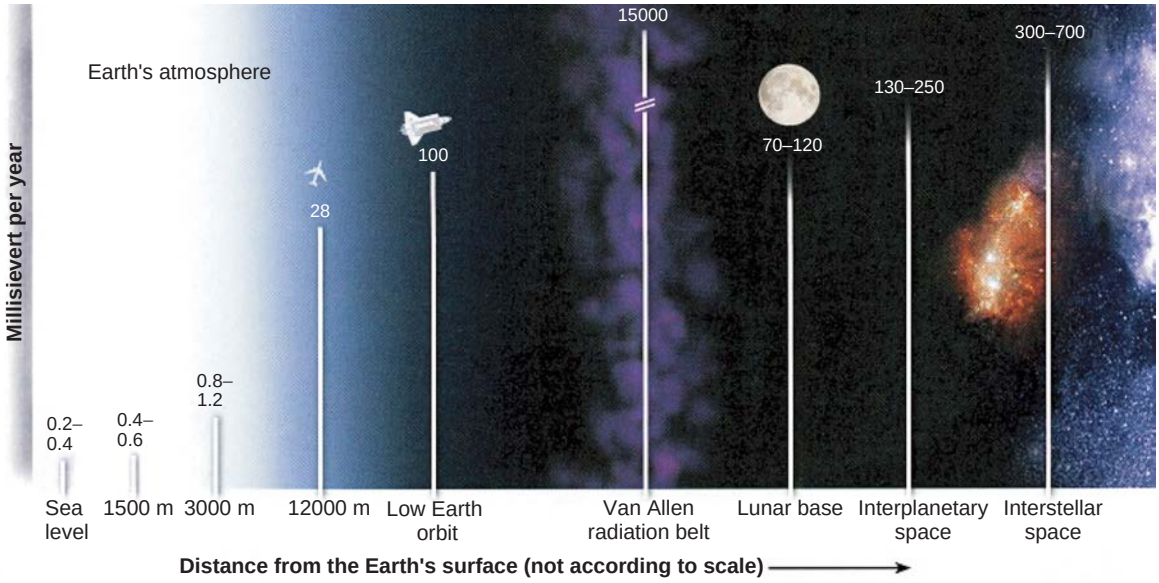


Figure 7.6.4: *Radiation exposition for different mission scenarios: LEO, interplanetary transit, Moon, Mars (Source: Kent Snodgrass).*

case (1 for gamma rays, approx. 20 for alpha particles, approx. 10 for high-energy ions) [7.6.1].

Figure 7.6.4 shows the supposed radiation contamination in millisieverts (mSv) for the various mission scenarios (LEO, interplanetary transit, Moon, Mars), whereby the maximal dosage per lifetime according to present knowledge should not exceed 250 mSv. It can be seen that radiation exposure in scenarios 2 to 4 is substantial and that measures to protect the space vehicle, lunar or Martian habitat must be taken [7.6.28], [7.6.29].

7.6.6 Psycho-Physiological Problems Arising from Residence in Space

7.6.6.1 Biorhythms

In space, humans are almost completely isolated from the physical and social environmental factors which normally determine their biorhythms. The terrestrial alternation between day and night is mimicked by controlling the artificial illumination in the space vehicle, and the daily fluctuations in ambient temperature are also artificially maintained, to mention only some of the relevant factors. **Social factors,** whose origins are family, workplace and friendships, are lacking. **Endogenous biorhythms** must adjust to a large extent to the mission's predetermined work and relaxation rhythms. The duration and quality of sleep have to be so optimized that performance readiness and a subjective feeling of well-being are maintained. However, the need to more or less stringently organize the **daily routine** means that individual needs have to be adapted to the needs of the entire group, and this for the duration of the entire mission. Negative consequences are inevitable because over long periods of time the astronaut's psychological state and work capacity fluctuate considerably. In this respect observations made during terrestrial isolation studies do not differ significantly from Russian experiences on long missions [7.6.4]. The daily schedule during the Russian isolation experiences is as follows: work period 09:00 to 17:00, resting 17:00–23:00, sleeping 23:00 to 07:00, getting-up phase 07:00–09:00, with meals four times a day. Presumably, there is accordingly little opportunity for an individual style which significantly varies from this routine.

Measurements of activity patterns with the help of accelerometers during a four-week ESA isolation study produced fairly similar results on workdays for all six test subjects [7.6.6]. Measurements on weekends showed, as expected, more individualized behavior patterns. Especially the distribution of sleep periods differed from the normal weekday pattern. However, strict regulation of the daily schedule should mean respecting natural sleep requirements, which during the day are especially pronounced after the main meal. It is known that forced, artificially extended periods of wakefulness negatively influence cognitive abilities, something which should definitely be avoided on long missions. But there is no general strategy which can be applied to the situation in space. The conditions under isolation and the length of the particular mission influence the sleep–wake rhythm as well as the cognition of the astronauts. Russian experience has shown that after an initial 14-day adjustment period a mission phase follows in which there is full adaptation until about the 90th day. Afterward, sleep problems can be expected, the activity of the crew members is reduced, and a reduction in the range of their interests can be observed. Increased nervousness and tiredness become evident. Since these phenomena were found during terrestrial isolation studies as well as in space, it is hard to identify the role played by weightlessness. It is currently unknown how **weightlessness** affects biorhythmic processes at the cellular and molecular levels, and what the consequences are for the entire system. In the future this will be a field for appropriate experiments [7.6.30].

7.6.6.2 Isolation and Constricted Movement

In the course of dealing with the physiological consequences of long periods in space, medical doctors had to consider the problems of isolation, solitude, loneliness and confinement of movement [7.6.31]. **Confinement** is often connected with a **lack of exercise,** which plays a role in the discussion of corrective measures (see Section 7.6.7 below) as well as in this section. But isolation and confinement are not problems limited to space flight only; they can be observed every day in society. In the large cities of Western Europe almost 50% of the population lives in single-person households, either by choice (singles, primarily young

Table 7.6.2: *Normal stimulus range on Earth.*

Stimulus parameter	Percentage of total spectrum (hypothetical)
1. Gravity	40% proprioreceptors (skin, muscles, joints)
2. Light	30% eyes
3. Sound	10% ears
4. Smell, taste	10% nose, mouth
5. Temperature	10% skin, CNS (Central Nervous System)

people) or because of external circumstances like the death of a partner. This trend is reinforced by increasing longevity. In order to identify the physiological effects of this change in ambient conditions, new, **noninvasive technologies** must be developed for **continuous data collection.** They should have low energy requirements, not get in the way when they are worn, and be easy to use by a layperson. This continuous flow of signals produces large amounts of data which can only be managed with the help of modern information technology. Time-series analyses are particularly valuable. When recording and processing data, the **spectrum of stimuli,** which is normal on Earth, should be kept in mind (Table 7.6.2).

For example, if gravity is lacking in space, then 40% of the body's stimuli disappears. In the case of the other parameters, noise, smell and taste, no large variations due to physical confinement are expected.

Living together in close quarters always means consideration for others; not what everybody wants, but what is the least disruptive has to be determined. This leads to an overall leveling down and to **monotony.** Isolation and confinement therefore mean a reduction or deficit in stimuli. In the best case this can lead to relaxation followed by recovery, particularly if one has been living in an overstimulated environment. If the removal of stimuli is unavoidable, such as for long residence in space, the first step is light **deconditioning** leading to a vicious circle as described in Figure 7.6.5.

In LEO this can be combated by the indicated countermeasures which provide psychological and social support (telephone calls to family members) [7.6.31]. Physical refreshment can for example be produced by thermal stimulation in an otherwise thermally monotonous environment. Such **stimulus substitution measures** are absolutely necessary in a

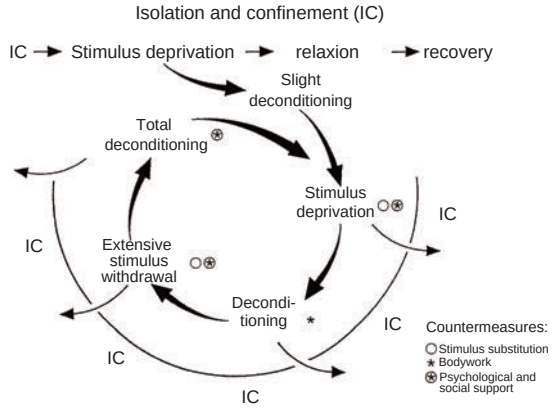


Figure 7.6.5: *Stimulus deprivation, stimulus withdrawal and the vicious circle which has to be overcome with appropriate corrective measures during long space flights (Source: [7.6.14]).*

monotonous environment in order to break out from the vicious circle which ends in total deconditioning. This situation can lead to not being able to readjust to terrestrial conditions. Stimulus substitution not only helps to make the stress associated with residence in space bearable, but also improves the chances for smooth readaptation to terrestrial conditions. Considering the experience gathered under space conditions, rehabilitation measures on Earth, such as after long immobility and sickness, need reevaluation.

The effects of a reduced **social spectrum** should not be underestimated. A crew consists of four to eight members at most, and they have to get along. The communication within such a group can vary considerably during a mission, as suggested in Figure 7.6.6.

The number of verbal communications among group members during the joint meal in a 28-day period of isolation was recorded. On the second day it can be noted that the communication of person D with commander C dominates; the other members are almost excluded. But by the end of the mission (26th day), person D is totally isolated during the meal. This social isolation within the group can have somatic effects which must be avoided [7.6.32]. One reason for isolation might be modern computer games, which each person can play alone in some corner of the space vehicle resulting in sitting back to back at leisure time without communicating. Competitive group games can counteract such a development. Crew selection is important here, so that the team

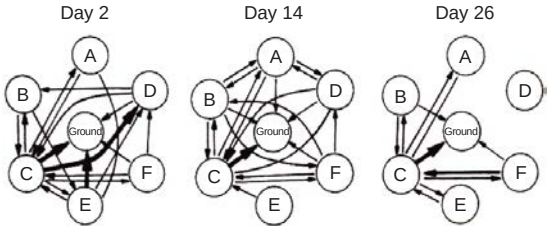


Figure 7.6.6: *Communication networks on days 2, 14 and 26 during team isolation (Source: modified from [7.6.32]).*

has the right composition. Only then can it fulfill its task in the long term.

It can be said in summary that a multitude of parameters affect the **group cohesion** and **performance** of a team: cultural background, the leadership qualities of the commander, sex, age, emotional stability, expertise and cooperativeness, to name but a few [7.6.3], [7.6.15]. All these factors must be considered when a crew is selected for a long stay in space. Physiological performance factors also play a role, but not necessarily a dominant one, once the basic parameters have been achieved (well trained at least to an average level). Psychologists have an important role in deciding what is the right data to consider.

7.6.7 Corrective Measures

It should first be pointed out that the physiological changes noted so far under weightlessness are **adaptations which are appropriate** for the space environment. The Russian cosmonaut Poljakov lived in the Russian MIR station for periods up to 14 months. Medical physiological problems usually occur when the astronauts are again subjected to a higher level of gravity, whether on Earth or in the future on the Moon or Mars. But here, particularly in the case of extravehicular activities in the context of a mission, physical fitness is the *conditio sine qua non*. Appropriate corrective measures must be taken, especially those assuring that the proper functioning of the **cardiovascular, thermoregulatory skeletal and muscle systems** is maintained. Providing emergency medical care for missions to the Moon or Mars could, just to consider the operational reasons alone, always only be possible to a limited extent. For presence in LEO it is another matter, since in a medical emergency an astronaut or cosmonaut can be evacuated relatively quickly to Earth (rescue capsule, emergency descent), something which is not possible for lunar and certainly not for Mars missions.

Among the **corrective measures** in the narrower sense can be counted all those treatments and therapeutic measures which serve to guarantee or restore the physical, mental and spiritual well-being and health of astronauts before, during and after a space flight. Physical, biochemical, chemical, pharmaceutical, biological, and psychological measures are possible. Table 7.6.3 provides an overview of various possibilities.

In the preparatory phase of a space flight they include **crew selection** ("select in") in order to assure a crew who are as efficient as possible and in which there is the lowest risk of interpersonal conflict. Physical and mental training programs have to be made available, and care must be taken to achieve the required physical fitness and high, continuous mental performance levels and alertness, not to mention programs which prepare the organism for altered circadian rhythms. During the flight most of the corrective measures concern providing and carrying out an **exercise program** of several hours with the help of customized training methods like treadmills, flywheels, rowing equipment, a "penguin suit," a vibration platform, etc. After return from space intensive psycho-physiological training programs are necessary to accustom the astronauts to the terrestrial environment and to their social milieu ("postflight disorders"). It has been shown that this **postflight phase** to reestablish the original preflight situation requires almost as much time as the space flight itself (Figure 7.6.7).

An exception in the physiological realm has to do with the changes which occur in **bone metabolism,** which according to the present state of scientific knowledge cannot be completely returned to the initial values.

7.6.7.1 Cardiovascular System

In order to avoid **orthostatic hypersensitivity** and **physical deconditioning** on the part of the astronauts

Table 7.6.3: *Measures to prevent deconditioning in space.*

Physical methods	Bodywork	Diet control	Pharmacological support	Environmental control
Controlling blood volume distribution	Avoiding muscle atrophy	Fluid balance		Stress avoidance, stress management
LBNP	Ergometer	Salt/water absorption	Digestive enzymes	Optimization of daily routine
g-suit	Treadmill	Energy uptake		
Blood pressure sleeve	"Penguin suit"		Pancreatic enzymes	Psychological support
Electrical stimulation	Expanders			

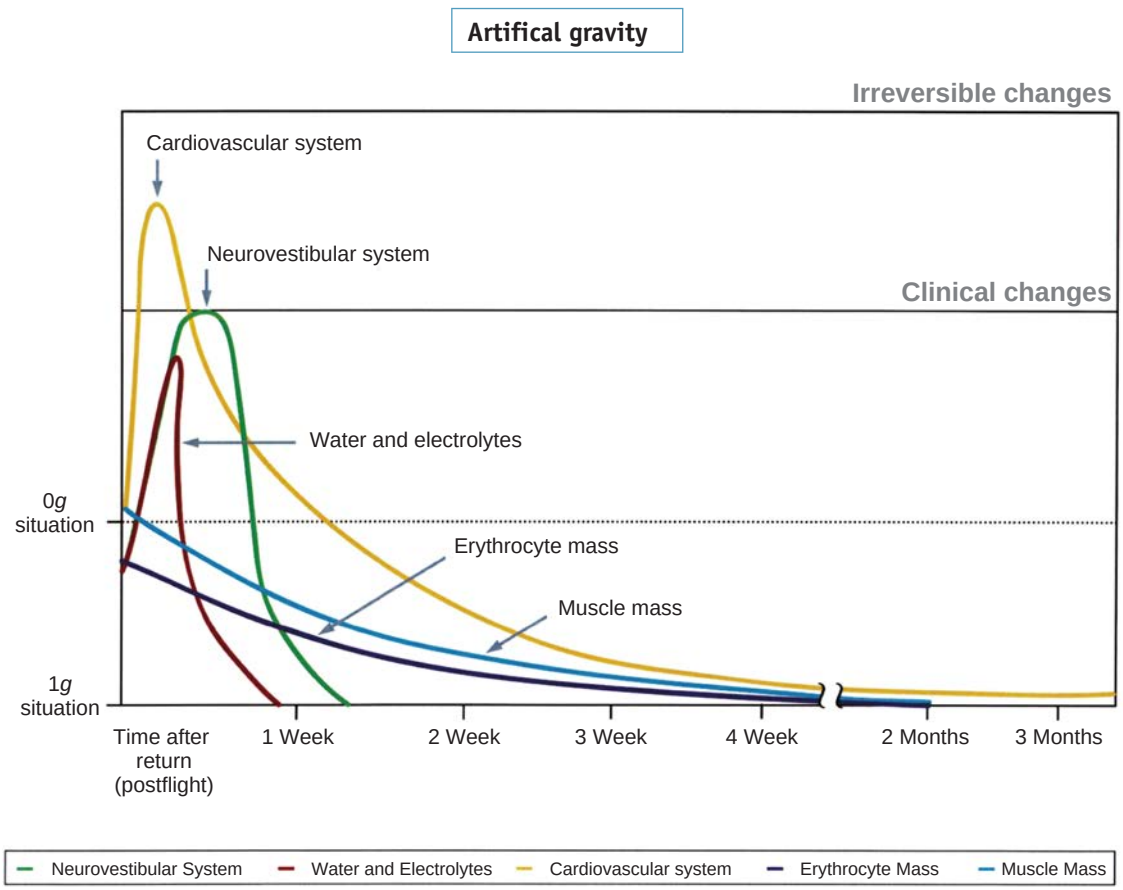


Figure 7.6.7: *Temporal progression of physiological changes during readjustment to conditions on Earth. Those changes in body systems are shown for which readjustment is not significantly dependent on the duration of presence in space. The adjustment curve of calcium metabolism is lacking since adequate data was not yet available (Source: [7.6.15]).*

during their presence in space [7.6.33], [7.6.34], the following should be available for daily training periods: bicycle ergometer, treadmill and rowing machine, among other equipment. This variety is necessary not only for physiological reasons, but also from a psychological point of view, particularly to avoid boredom. During the training units it should be assured that basic cardiovascular and pulmonary data can be noninvasively monitored, stored and analyzed (V_{o_2max}, heart frequency, blood pressure,

breathing rate, skin and body basal temperature, oxygen saturation, etc.). Before landing, there should be a possibility for the astronauts to be supplied with liquid in order to increase their **plasma volume.** This kind of liquid replenishment can still be carried out after **orthostatic intolerance** has developed. For this treatment practical devices for intravenous infusions and increasing plasma volume must be at hand for emergencies and the astronauts have to be trained in the technique of intravenous infusion. During the critical phase of reentry, **antigravity suits** are worn, which support blood flow toward the upper body by applying pressure to the legs; in addition, horizontally oriented seats are used [7.6.15]. Despite these countermeasures orthostatic intolerance remains for the most part an unsolved clinical and mission-relevant problem.

7.6.7.2 Muscle and Skeletal System

Power training to maintain body strength and bone mass is the second pillar of physical exercise, the immediate motivation being here, too, to preserve the strength of the hands for work outside the space vehicle [7.6.3], [7.6.15], [7.6.24], [7.6.35]. In addition to various elastic band and stretching systems for resistance training, the procedures mentioned above for cardiovascular training can also be used, such as rowing machines or a treadmill, in which the trainee is pulled toward the machine via a system of elastic resistance bands (high-impact training to maintain bone structure). It has been shown that already a slight amount of **vibration** exerts sufficient stimuli to maintain bone mass. The application of vibrations during the flight could help shorten the current tedious exercise regime. Pharmacological approaches to maintain muscle tone and the skeletal system include doses of amino acids to stimulate protein synthesis and of biphosphonates to reduce global calcium loss and maintain bone density.

7.6.7.3 Neurosensory System

The lack of gravitational stimulation in space leads to incompatible information coming from the visual, vestibular and kinesthetic system, which can lead to what is known as **space motion sickness** [7.6.15], [7.6.24]. Although symptoms like dizziness, nausea and feeling ill recede after a few days, they often return upon entry into a gravitational field, which could hinder the continuation of the mission once the astronauts arrive on Mars. Unfortunately, there are large differences in the nature and particular expression of symptoms, both between **individuals and for the same person on different occasions.** These often do not agree with observations made in experiments on Earth, making it difficult to predict susceptibility to space motion sickness. Selecting individuals with a high tolerance for movement, acceleration and deceleration did not produce the desired results. The use of biofeedback and autogenous training was successful in reducing some of the symptoms. **Medication** (e.g., promethazine) can also reduce the severity of space motion sickness, but it also can lead to a drop in performance (affecting reaction time and pattern recognition, for example) and can negatively influence sleep and overall mood.

7.6.7.4 Radiation Protection

Selection of crew members based on the lowest possible number of changes in particular gene loci known to be oncogenes, as well as the extraction and storage of **bone marrow** as a regeneration possibility, can be carried out also before the flight as precautionary measures to avoid later malignant illness. **Pharmacological radiation protection** is available in the form of free radical scavengers as well as innovative radioprotective substances (like DNA-binding WR-33278). Protease inhibitors and substances like dimethylsulfoxide (DMSO) can hinder the evolution of an already radiation-damaged cell into a cancerous cell cluster. This development gives rise to the hope that pharmacological approaches can be developed to counter radiation risks.

On the technical side, the **hull** of the space vehicle itself supplies the most protection. Since most of the risk comes from high-energy solar particles moving slower than the speed of light, a **satellite-supported early warning system** could be put into operation. As soon as a solar event (eruption, solar wind) has been identified, the information can be relayed to the crew on-board the space vehicle, which would give them enough time to get to an especially shielded area [7.6.28], [7.6.29].

An innovative technical approach would be to use a **magnetic deflector** to divert solar particles from the space vehicle. New superconducting electromagnetic coils which no longer require supercooling have shown promising results and are financially affordable. But data is not yet available on the long-term effects of comparable magnetic fields on human tissues.

7.6.7.5 Personal Databases for Astronauts

It can been seen from all that has been said above that there are great individual differences in the progression and extent of adjustments which astronauts make to their new environment (e.g., as to the neurovestibular system, liquids and electrolytes, cardiovascular system, red blood cells, bone and calcium metabolism, nonfat body mass) [7.6.3], [7.6.15]. It is essential that these changes be recorded and monitored for each individual astronaut as completely and as long as possible before (baseline data), during and after their presence in space. This data is, for example, essential for establishing guidelines for exercise programs designed to counteract cardiovascular deconditioning. Such data is also needed in case of emergencies.

7.6.8 Outlook

Space medicine research is characterized by innovative scientific approaches and pioneering technological applications. Although an apparently very exotic specialty, it has a frequently underestimated high **clinical relevance.** Specifically, practical questions of high-altitude, climate, diving, sport and occupational medicine, and also rehabilitation and even isolation research, are subjects of interest (osteoporosis, cardiovascular illness, corrective measures and exercise regimes, etc.). Space medicine research thus extends far beyond the narrow realm of space physiology and space medicine. **Terrestrial simulation studies** with their customized and sometimes unique simulation models and equipment (bed rest, lowered head positions, isolation and immersion) are superbly suited for addressing interdisciplinary human physiology concerns in great scientific depth and breadth. At the same time, such studies serve to introduce the **next generation of scientists** to this field of investigation and to interest them in the subject, in order to ensure that research continues. This problem should not be underestimated in the typically extremely long planning periods of space activities. For scientific, technological and economic reasons, research in space medicine and space physiology must be regarded as an integral, essential component of space policy also in the future.

Bibliography

[7.6.1] Messerschmid, E., Bertrand, R. *Space Stations.* Heidelberg: Springer Verlag, 1999.

[7.6.2] Harland, D.M. The MIR Space Station. In Mason, J. (ed.), *The Story of Space Station MIR*, Wiley-Praxis Series in Space Science and Technology. Chichester: John Wiley & Sons, Ltd, 1997.

[7.6.3] Clément, G. *Fundamentals of Space Medicine*, Space Technology Library. Heidelberg: Springer Verlag and Microcosm Press, 2001.

[7.6.4] Gushin, V.I., Kholin, S.F., Ivanovsky, Y.R. Soviet Psychophysiological Investigations of Simulated Isolation: Some Results and Prospects. In Bonting, S.L. (ed.), *Space Biology and Medicine.* Greenwich, CT: JAI Press, 1993.

[7.6.5] Maillet, A., Normand, S., Gunga, H.-C. *et al.* Hormonal, Water Balance, and Electrolyte Changes During Sixty-Day Confinement. In Bonting, S.L. (ed.), *Space Biology and Medicine.* Greenwich, CT: JAI Press, 1993.

[7.6.6] Tobler, I., Borbély, A. Twenty-Four Hour Rhythm of Rest/Activity and Sleep/Wakefulness: Comparison of Subjective and Objective Measures. In Bonting, S.L. (ed.), *Space Biology and Medicine.* Greenwich, CT: JAI Press, 1993.

[7.6.7] Comet, B. Limiting Factors for Human Health and Performance: Microgravity and Reduced Gravity. In Study on the survivability and adaptation of humans to long-duration interplanetary and planetary environments, Technical Note 2: Critical assessments of the limiting factors for human health and performance and recommendation of countermeasures. *HUMEX-TN-002*, 2001.

[7.6.8] Heer, M., Baisch, F., Kropp, J. *et al.* High Dietary Sodium Chloride Consumption May Not Induce Body Fluid Retention in Humans. *Am. J. Physiol. Renal Physiol.*, **278**, 585–595, 2004.

[7.6.9] Titze, J., Maillet, A., Lang, R. *et al.* Long-Term Sodium Balance in Humans in a Terrestrial Space Station Simulation Study. *Am. J. Kidney Dis.*, **40**, 508–516, 2002.

[7.6.10] Blüm, V. Künstliche Ökosysteme und Raumfahrt. In Rahmann, H., Kirsch, K. (eds.), *Mensch – Leben – Schwerkraft – Kosmos.* Stuttgart: Günter Heimbach Verlag, 2001.

[7.6.11] Billica, R.D., Simmons, S.C., Mathes, K.L. *et al.* Perception of the Medical Risk of Spaceflight. *Aviat., Space Environ. Med.*, **67**, 467–473, 1996.
[7.6.12] Widing W., Halliwell J.,R., Karemaker I.M.. Orthostatic Intolerance after Space Flight. *Acta Physiol. Scand.*, **144**, 604, 1992.
[7.6.13] Morey-Holton, E.R., Whalen, R.T., Arnaud, S.B. *et al.* The Skeleton and its Adaptation to Gravity. In Fregly, M.J., Blatteis, C.M. (eds.), *Handbook of Physiology: Section 4, Environmental Physiology*, Vol. I. New York: Oxford University Press, 1996.
[7.6.14] Gunga, H.-C., Kirsch, K., Röcker, L. *et al.* Body Weight and Body Composition During Sixty Days of Isolation. In Bonting, S.L. (ed.), *Space Biology and Medicine*. Greenwich, CT: JAI Press, 1993.
[7.6.15] Nicogossian, A., Leach Huntoon, C., Pool, S.L. *Space Physiology and Medicine*. Second Edition. Philadelphia: Lea & Febiger, 1989.
[7.6.16] Blomquist, C.G. *SLS-1 Final Report: Cardiovascular Adaptation to Zero Gravity*. Dallas, TX: University of Texas Southwestern Medical Center, 1994.
[7.6.17] Thornton, W.E., Hedge, V., Coleman, E. *et al.* Changes in Leg Volume During Microgravity Simulation. *Aviat., Space Environ. Med.*, **63**, 789–794, 1992.
[7.6.18] Smith, J.J., Ebert, T. General Response to Orthostatic Stress. In Smith, J.J. (ed.), *Circulatory Response to the Upright Posture*. Boca Raton, FL: CRC Press, 1990.
[7.6.19] Watenpaugh, D.E., Hargens, A.R. The Cardiovascular System in Microgravity. In Fregly, M.J., Blatteis, C.M. (eds.), *Handbook of Physiology: Section 4, Environmental Physiology*, Vol. I. New York: Oxford University Press, 1996.
[7.6.20] Hinghofer-Szalkay, H.C., Moore, D. Some Comments on Biological Aspects of Life Support Systems. In Moore, D., Bie, P., Oser, H. (eds.), *Biological and Medical Research in Space*, ESA Publication. Heidelberg: Springer Verlag, 1996.
[7.6.21] Aubert, A.E., Beckers, F., Verheyden, B. Cardiovascular Function and Basics of Physiology in Microgravity. *Acta Cardiol.*, **60** (2), 129–151, 2005.
[7.6.22] Alfrey, C.P., Udden, M.M., Leach-Huntoon, C. *et al.* Control of Red Blood Cell Mass in Spaceflight. *J. Appl. Physiol.*, **81**, 98–104, 1996.
[7.6.23] Gunga, H.-C., Kirsch, K., Baartz, F. *et al.* Erythropoietin under Real and Simulated Microgravity Conditions in Humans. *J. Appl. Physiol.*, **81**, 761–773, 1996.
[7.6.24] Edgerton v., R., Roy, R.R.: Neuromuscular Adaptations to Actual and Simulated Spaceflight. In Fregly, M.J., Blatteis, C.M. (eds.), *Handbook of Physiology: Section 4, Environmental Physiology*, Vol. I. New York: Oxford University Press, 1996.
[7.6.25] *Süddsche Z.*, No. 271, p. 13, November 24, 2006.
[7.6.26] Stein, T.P., Leskiw, R.J., Schluter, M.D. *et al.* Energy Expenditure and Balance During Spaceflight on the Space Shuttle. *Am. J. Physiol. Regul. Integr. Comput. Physiol.*, **276** (6), R1739–R1748, 1999.
[7.6.27] Stein, T.P., Leskiw, R.J., Schulter, M.D. *et al.* Protein Kinetics During and After Long-Duration Spaceflight on MIR. *Am. J. Physiol. Endocrinol. Metab.*, **276** (6), E1014–E1021, 1999.
[7.6.28] Nelsen, G.A. Radiation in Microgravity. In Fregly, M.J., Blatteis, C.M. (eds.), *Handbook of Physiology: Section 4, Environmental Physiology*, Vol. II. New York: Oxford University Press, 1996.
[7.6.29] Kiefer, J., Schenk-Meuser, K., Kost, M. Radiation Biology. In Moore, D., Bie, P., Oser, H. (eds.), *Biological and Medical Research in Space,* ESA publication, Heidelberg: Springer Verlag, 1996.
[7.6.30] Moore, D., Cogoli, A. Gravitational and Space Biology at the Cellular Level. In Moore, D., Bie, P., Oser, H. (eds.), *Biological and Medical Research in Space*, ESA publication. Heidelberg: Springer Verlag, 1996.
[7.6.31] Holland, A.W. NASA Investigations of Isolated and Confined Environment. In Bonting, S.L. (ed.), *Space Biology and Medicine*. Greenwich, CT: JAI Press, 1993.
[7.6.32] Bergan, T., Sandal, G., Warncke, M. *et al.* Group Functioning and Communication. In Bonting, S.L. (ed.), *Space Biology and Medicine*. Greenwich, CT: JAI Press, 1993.
[7.6.33] Fortney, S.M., Schneider, V.S., Greenleaf, J.E. The Physiology of Bed Rest. In Fregly, M.J., Blatteis, C.M. (eds.), *Handbook of Physiology: Section 4, Environmental Physiology*, Vol. II. New York: Oxford University Press, 1996.
[7.6.34] Frey, M.A.B., Charles, J.B., Houston D.E. Weightlessness and Response to Orthostatic Stress. In Smith, J.J. (ed.), *Circulatory Response to the Upright Posture*. Boca Raton, FL: CRC Press, 1990.
[7.6.35] Oganov, V.S., Schneider, V. Skeletal System. *Space Biol. Med.*, **3,** 247–266, 1996.

7.7 New Technologies and Robotics

Gerd Hirzinger, Klaus Landzettel and Clemens Kaiser

Technology missions are undertaken to test and further develop the space technologies themselves as well as to qualify them for subsequent scientific or commercial applications. Such missions usually involve challenging new concepts, and mission preparation frequently has an additional positive **spin-off effect** on developments not related to space activities.

Examples of missions designed to investigate satellite technologies are given elsewhere, particularly in Section 8.5 on microsatellites. Thus the BIRD satellite is being used to test a variety of new approaches in the fields of infrared sensors, attitude regulation, automatic navigation and on-board software. The first Galileo navigation satellites, GIOVE A and B, are also in this category.

It is not possible to present a complete overview of all technology missions, but some insights into typical aspects of the technologies involved can be gained from a look at some selected examples. This section features two applications, **space robotics** (Section 7.7.1) and **on-orbit servicing** (Section 7.7.2). Both areas have such high relevance for future space missions that they are worth closer inspection.

7.7.1 Space Robotics

Transporting persons into space, keeping them alive there and bringing them safely back to Earth requires a tremendous effort, which also finds expression in the high cost of this endeavor. The explanation is that humankind evolved in the course of some hundred thousand years to be optimally suited to conditions on Earth. For the foreseeable future people will not be able to accustom themselves to the hostile conditions of space – the consequence being the enormous effort to create appropriate **life support systems**, including expensive space suits.

Robotic systems are another matter. The field itself is little more than 40 years old and can be adapted within a short time and at relatively low cost to the space environment. When combined with powerful remote control and telepresence technology, the sometimes overrated question of how long it will take for robots to become really intelligent and able to work autonomously becomes fairly irrelevant. Once the basic concepts behind "an extended arm in space" have been understood and realized, then one can switch at will during a mission among all possible degrees of **shared autonomy** between an operator on the ground and a robot in space, incorporating at each point in time the momentary state of the art in achieving the highest possible level of autonomy on-board a spacecraft. This kind of activity in space with remote control from Earth, coupled with the rapid advances in robot technology, offer an increasingly cost-effective alternative to human space flight. But this option is also becoming more attractive because so-called **telepresence** concepts are making it possible to give many people on the ground a realistic feeling of how it would be if they were "on location" in space; up until now this feeling was the preserve of only a few astronauts and cosmonauts.

The questions naturally arise of what robots should achieve in space and which control technology is most suited to the purpose.

In a space system, as in any technical system, problems can arise which threaten the system itself, its environs, the astronauts on the ISS, or the population on Earth. Those problems will require **repairs** accomplished with human intervention. In the case of the ISS it is the astronauts themselves who must attempt to solve any problems, possibly during extravehicular activities, which involve dexterity, intensive prior training and enormous physical effort. The idea of sending astronauts to defective satellites in low Earth orbit (LEO) is even more difficult and expensive – the last planned Hubble repair was long threatened by the high cost and risks connected with the Space Shuttle. So far, space systems located above about 500 km cannot be reached by astronauts on the shuttle and are thus irreparable, when telecommands from ground and software modifications fail to solve the problem.

Satellites equipped with robotic arms, so-called **robonauts**, are controlled from Earth and in the future should be in a position to solve many problems. Remotely controlled robonauts will support astronauts in routine and maintenance activities on the space station, retrieve and repair satellites which are tumbling uncontrolled, and make space available in the popular geostationary orbit by removing redundant spacecraft.

So far there has been no possibility to service space systems in the economically most efficient way by robots; for example, to have remotely controlled robots capture an object and cause it to fall into the Atlantic or Pacific Ocean, thereby avoiding problems with debris (see also Section 7.7.2). In 2003 the European Parliament challenged Europe to assume international leadership in this area. The term **exploration**, by

contrast, refers to planetary and solar research. In this field robotic systems are an even more cost-effective alternative to human space flight.

Either autonomously or controlled from Earth, rovers equipped with robotic arms can collect and analyze rock samples, or make preparations for human residence on distant planets. How **remote control** functions depends on what **communications links** are available and on the signal travel times. A robot on Mars would have to function as autonomously as possible. The operator on the ground can only define the goal, leaving it to the system in space to independently determine via calculations how it is to be accomplished. Here, direct remote control is not possible because of the long signal propagation delay (about 15 minutes). Such high signal propagation delays mandate that the robotic system independently navigates, recognizes objects, and knows how to grasp and manipulate them.

On the other hand, in the case of communication links with relatively short signal propagation delays (under 0.5 seconds), use can be made of the **telepresence control technologies** mentioned above. This very powerful operating tool puts the operators at the ground station in a position to use the information supplied, such as stereo images, forces and moments, to react as if they were themselves in space.

Fortunately, it is possible for all robotic systems orbiting the Earth to comply with the "half-second control loop boundary limit", given adequate (although not yet generally available) **communications infrastructure**. This limit primarily goes into effect if **haptic feedback** is desired; in other words, sensing and selectively exerting forces in space. In the **visual channel**, however, people are fairly well able to manage delay times up to 1 second. If the robotic system is in LEO and if one is not satisfied with the short ground station pass durations of typically 7–10 minutes (whereby the signal propagation delay is then about 10 ms), then one requires a link to a geostationary satellite, from there to a robotic satellite in a low orbit (e.g., 300 km) and from there via an automatically tracking antenna back to the geostationary satellite and to the ground station. The control loop time is then in the realm of the required half second and contact times of 40 minutes are not untypical in such cases.

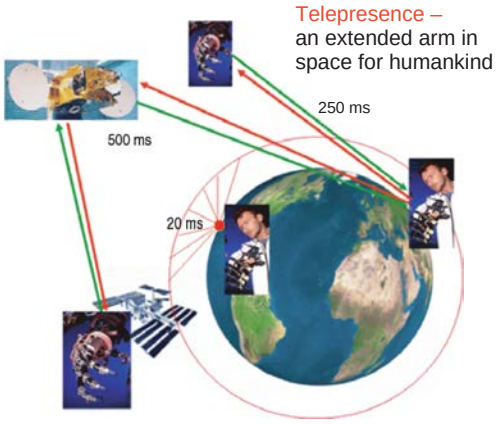


Figure 7.7.1: *CX-OLEV: the minimal "round-trip signal propagation time" to robots in LEO is 0.5 s, and to robots in GEO 0.25 s.*

Satellite repair and maintenance in geostationary orbit are even easier to accomplish with such a powerful telepresence operating mode because the signal propagation delay amounts to only about 0.25 seconds and the link is continuous (Figure 7.7.1).

7.7.1.1 Manipulators in Earth Orbit

The artificial arms which have been employed operationally in space are usually called manipulators and resemble cranes that are remotely controlled by the astronauts over a distance of a few meters (hence the designation **telemanipulators**). The first spectacular and successful manipulator in space was the **Shuttle Remote Manipulator System** (SRMS), a Canadian development which flew in space for the first time in 1981 in the cargo bay of the shuttle. This approx. 15 m long arm still often accompanies the shuttle; it has been used over 100 times so far, frequently to deploy and retrieve satellites and systems in need of repair (as in the case of the Hubble Space Telescope), but also to move astronauts around in free space (Figure 7.7.2).

Its six-jointed arm is practically without any sensor-supported autonomy and is controlled exclusively by the astronauts using two joysticks, one for translation and one for rotation of the end effector. This **controlling** requires extensive training because the long arm swings back and forth when it is moved, whereby the causative elasticities are primarily concentrated in the



Figure 7.7.2: *The shuttle arm transporting an astronaut (Source: NASA).*

joints. After the *Columbia* shuttle accident the arm was equipped with instruments which make it possible to inspect the shuttle for external damage. The successor of this arm is the so-called Canadarm II on the ISS, also known as the Space Station Remote Manipulator System (SSRMS).

An approx. 17 m long main arm is supplemented by two smaller arms of 3 m in length (**Special Purpose Dexterous Manipulator** (SPDM)), see Figure 7.7.3. The main arm can be transported on a platform that moves along rails extending the length of the space station. And since it has a completely symmetrical construction, it can also "walk" end over end along the station by repeatedly turning its "foot" into a "hand." This arm is also primarily intended to lift heavy loads, for example to exchange the storage batteries for the solar generators. One of its spectacular operations was to remove a heavy airlock from the shuttle's cargo bay and fasten it to the space station (Figure 7.7.4). Two astronauts were at the joysticks for six hours – actually consuming valuable astronaut time, because such operations can be perfectly well planned on Earth and remotely controlled or programmed. This has been convincingly demonstrated in the space robotic experiments described below.



Figure 7.7.3: *The multiconfiguration robotic system developed in Canada for the ISS (Source: CSA).*

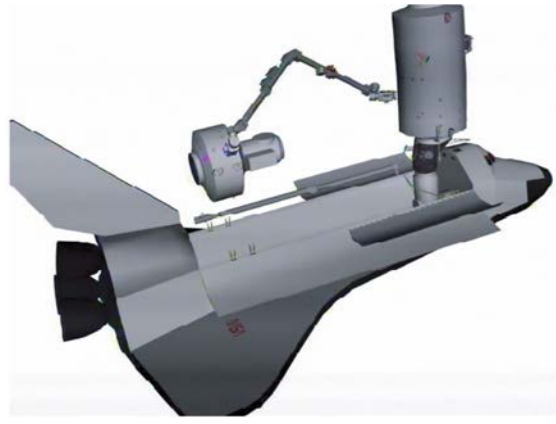


Figure 7.7.4: *Installing the airlock with the assistance of the ISS arm (Source: DLR).*

7



Figure 7.7.5: *Partial view of the space station with the ERA arm (Source: ESA).*

Canadarm is planned to be complemented on the space station by the European, approx. 11 m long ERA arm on the Russian portion of the space station (Figure 7.7.5, top) and by the Japanese JEM-SRMS arm of approx. 10 m length on the Japanese experimental module (JEM). The latter will have a small, 2 m long arm at its end, so that the entire system will have 2 × 6 or 12 degrees of freedom (the associated robotics term is **macro–micromanipulation**).

7.7.1.2 First Technology Experiments Leading to Semiautomatic Service Robots in Space

In the 1980s NASA set in motion the development of a **partially autonomous** free-flying three-arm robotic system for such purposes as servicing microgravity experiments (Figure 7.7.6).

With hindsight it can be said that this effort was premature, and after a few years NASA accordingly discontinued this spectacular development of a flight telerobotic servicer.

The first **partially autonomous space robot** remotely controlled from Earth, **ROTEX**, was successfully tested by DLR in 1993 as one of the experiments on the Spacelab D2 mission (Figure 7.7.7). The small ROTEX arm with its six joints was contained in a Spacelab rack and carried out prototype operations such as:

- Opening and reconnecting bayonet connectors.
- Assembling cube-shaped lattices.
- Capturing a small, free-flying object (a cube with flattened corners).



Figure 7.7.6: *Artist's impression of NASA's Flight Telerobotics Servicer (FTS), a three-arm robotic system (Source: NASA).*

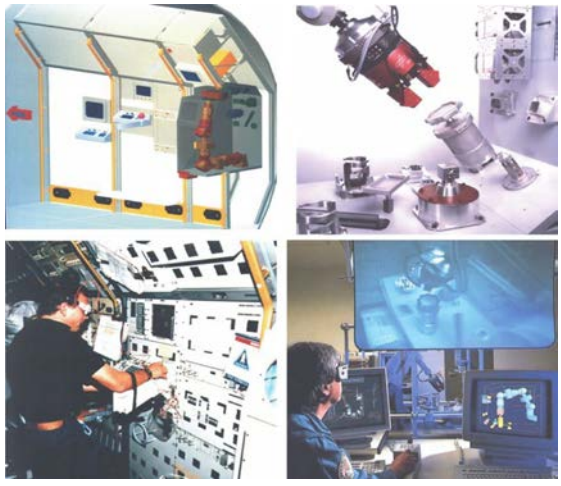


Figure 7.7.7: *ROTEX (Source: DLR).*

A number of operating modes still relevant today were demonstrated:

- **Direct remote control** with a trackball (which later evolved into the mass-produced space mouse); because of the long signal propagation delay of 5–7 seconds (back and forth to geostationary relay satellites), predicative 3D computer graphics were required.
- **Shared control** (shared autonomy), which means that the operator's instructions from Earth are interpreted only as general commands (also in the case of direct control) which are independently

refined by the robot using its own local sensors. The prerequisite was multisensory grasping technology; the two-point grasping element contained a total of 16 sensors, included two complementary moment-of-force sensor concepts, nine laser rangefinders, tactile sensors and stereo cameras.

- **Task-level remote programming** and reprogramming from Earth, with the task subsequently executed on-board at specified times. Here, too, multisensory "intelligence" and visualization of the task in a virtual world played an important role.

The most spectacular ROTEX experiment was certainly the automatic capture of a cube tumbling in a microgravity environment, because the images from the grasping had to be sent to the ground where a motion model was calculated (with a forecast for the ensuing 6 seconds) and, based on that result, the precalculated grasping element commands were relayed to the spacecraft (Figure 7.7.8).

Shortly thereafter, experiments began at DLR to study the **dynamic interaction** of a robot arm installed on a platform (Figure 7.7.9, bottom), which had the task of capturing another body free floating in microgravity (e.g., a satellite requiring repair). This situation was simulated in the laboratory with two industrial robots. It then became evident that the large telecommunication satellites had nothing which

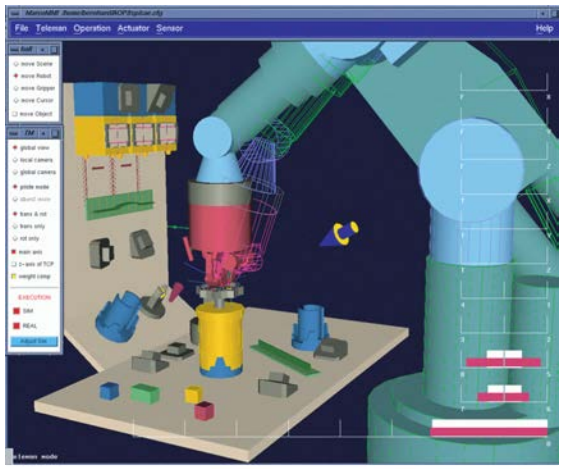


Figure 7.7.8: *Predictive (anticipative) computer simulation.*

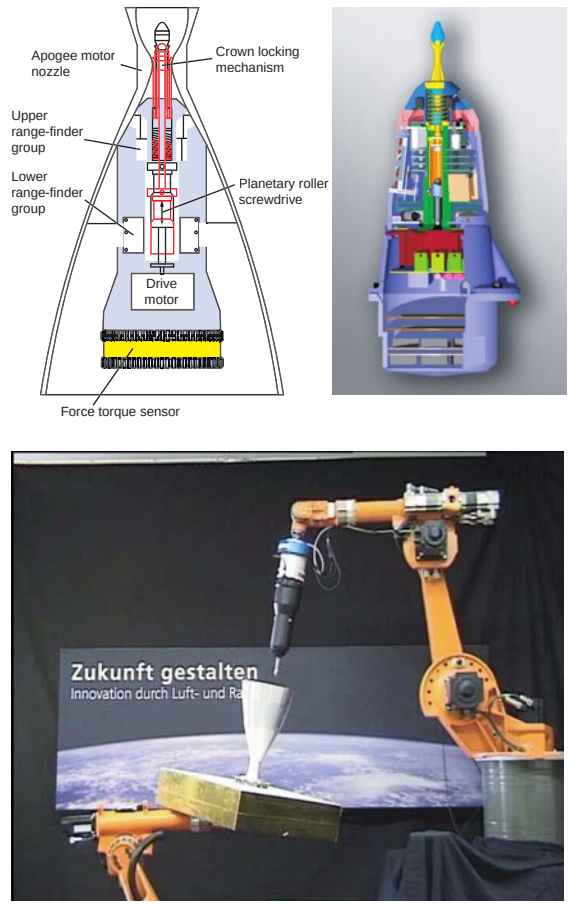


Figure 7.7.9: *The laboratory simulator showing the capturing technology (below) and the original capture tool (above left) as well as the current version (above right) (Source: DLR).*

7

could be used to grasp them with, except for their so-called apogee motor, which is the control nozzle that puts them into a geostationary orbit. This was the birth of the concept of a **capture tool** which, with the help of small cameras, radial laser rangefinders and force sensors, can enter into the nozzle and then activate a spreading and locking mechanism at the end of the capture tool, which makes it possible to securely grasp the satellite (see Figure 7.7.9, top, and also Section 7.7.2).

After this initial maneuver the "repair robot" can hold onto the satellite using a simple clamping mechanism and with its then freed arm fetch a tool to carry out the repair.

The TV-SAT 1 mission at the end of the 1980s, which failed because a solar panel did not deploy, provided the impetus to develop this capture tool. The docking of a servicing robot followed by the cutting of the solar panel clamp mechanism with metal shears would have made the satellite functional. Although shears to cut through the titanium clamps which kept the solar panels from unfolding had already been developed, a repair mission did not materialize. At the end of 2006 practically the same problem occurred with the initially highly celebrated Chinese telecommunications satellite Xinnuo 2. Its panels did not fully deploy either; the satellite had to be abandoned. Once again, the aforementioned shears would have sufficed to carry out the repair – another example of the importance of orbital robotic servicing from Earth.

As indicated, for such a capture it is crucial that the **reaction effects** which a moving arm exerts on its platform are kept under control, either by calculating them in advance or by compensating for them with reaction wheels (for the rotation) or cold gas jets (for the translation). A number of experts in robot dynamics are still working today on the **dynamic effects** of free-flying robots, including so-called null space motion in cases of kinematic redundancy (seven or more displacement degrees of freedom). Figure 7.7.10 shows that a seven-jointed robot which always traces exactly the same circular motion in space nevertheless displaces its platform differently depending on how its individual joints move (which in the case of the mentioned redundancy cannot be unambiguously determined).

An opportunity to investigate dynamic interactions in space was soon offered. Japan's space agency NASDA (now JAXA) launched the **first free-flying space robot** in 1997 as part of the ETS-VII mission (Figure 7.7.11). "Standing" on a platform it functioned without problems for over two years. As part of a German–Japanese joint project, it was remotely controlled from Tsukuba (Figure 7.7.12), also by German research partners under the leadership of the DLR institute.

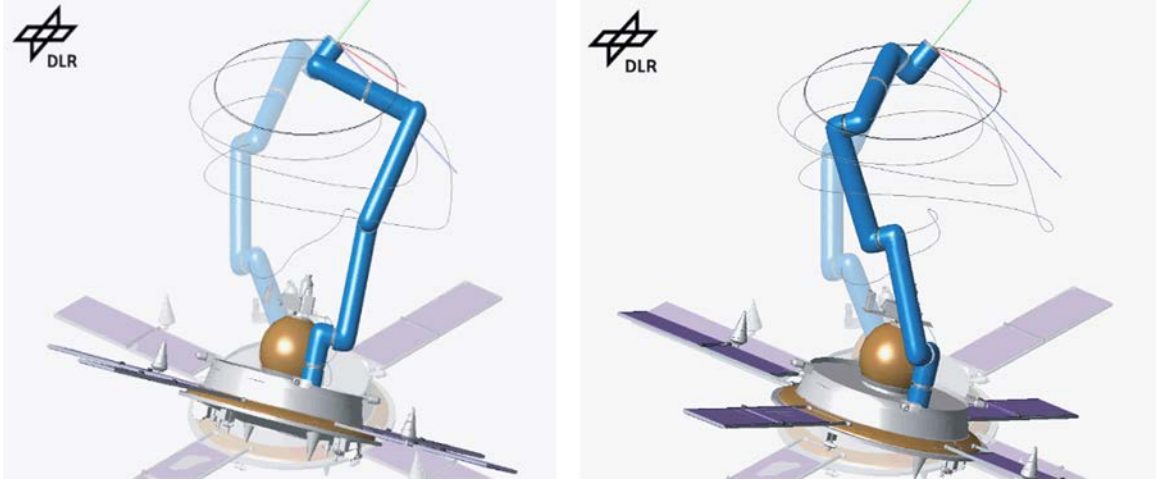


Figure 7.7.10: *Although executing the same movement, a kinematically redundant space robot can displace its platform in different ways (Source: DLR).*

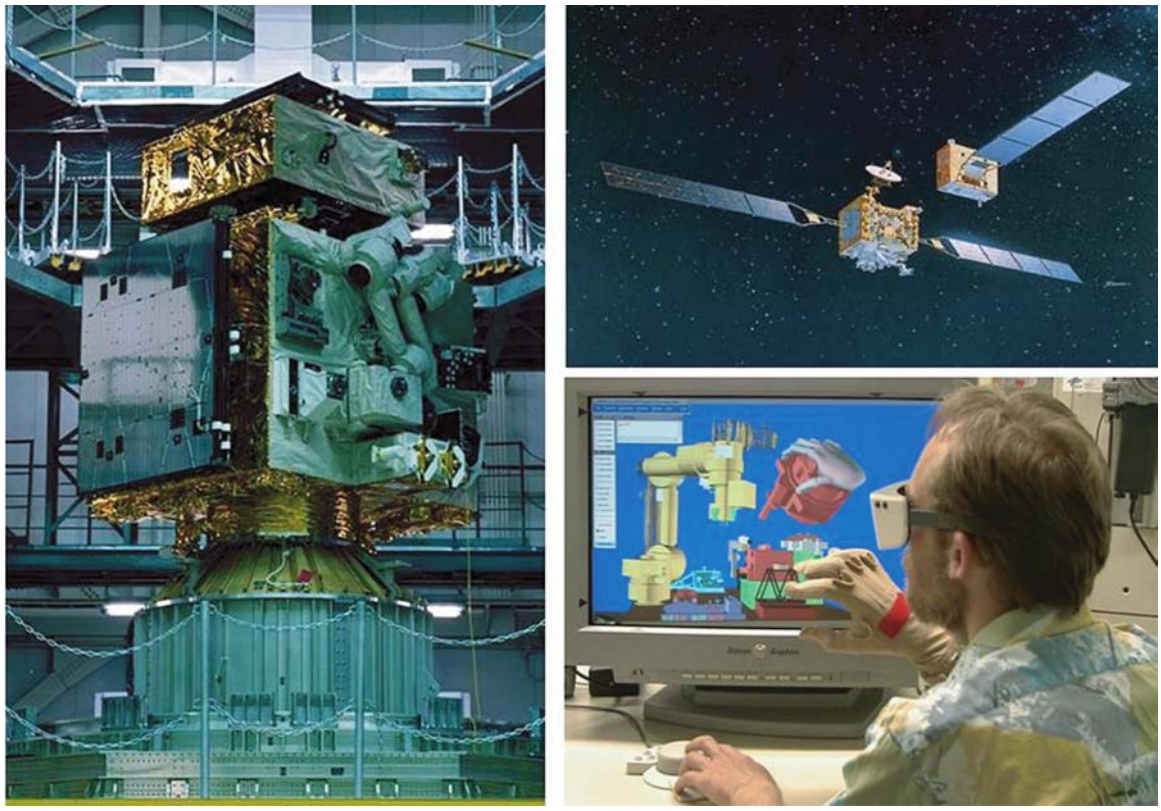


Figure 7.7.11: *ETS-VII, the first free-flying robot in space, with GETEX as the German contribution (Source: JAXA/DLR).*

This robot and its platform then had the task of docking with a target satellite launched at the same time and either on their own task board or on that of the target satellite carry out various assembly tasks and similar activities. As part of the joint project GETEX (with German participation involving DLR and Dortmund University, see Figure 7.7.11) two main goals were pursued:

- Demonstration of sensor-supported, **remote programming**, which had been considerably improved compared to ROTEX and included video.
- Control of **dynamic interaction**; the attitude regulation based on reaction wheels was turned off and the satellite was intentionally misdirected with arm movements (so-called swimming motions). It required two years of intensive research to clarify the discrepancy between predicted and measured satellite displacement. Factors such as inertial parameters (which were not precisely known because information about the exact amount of remaining fuel and its location in the tank was lacking) can in the meantime be estimated during the flight, quasi online.

7.7.1.3 Recent Technological Developments and Experiments

The remaining technology gaps were closed at the beginning of 2005 with the installation of the **ROKVISS experiment** on the outside of the space station. ROKVISS (Robot Component Verification on the International Space Station) provided the space flight qualification of DLR's torque-controlled lightweight robot joints on the one hand, and on the other hand demonstrated for the first time the performance of so-called telepresence control technology in



Figure 7.7.12: *Images from the DLR telerobotics station in Tsukuba (below left) and from space (Source: DLR).*

actual mission operation. Both aspects are described in somewhat more detail below.

With space flight as the motivation for **lightweight construction and energy efficiency**, the new robodrive propulsion motor was developed with its dramatically reduced weight and significant reduction of dissipation losses. When combined with innovative torque sensors in each joint, what are probably the most advanced robot arms to date were developed. Because of their sensitive "soft robotics" qualities and kinematic redundancy (seven joints) they are especially well suited for direct interaction with humans and can accordingly be regarded as the foundation for future "robonauts" (Figure 7.7.13). They are the first electromechanical arms which can carry considerably more than their own weight (13 kg). Maximum loads are not the primary concern in space, but weight is, because of launch costs, as are the forces exerted, because of the acceleration of load masses.

The joint units of ROKVIS were modified to withstand actual space conditions. Indeed, the conditions of the **space environment**, like solar irradiance on the one hand and the frigid temperatures of space on the other hand, high bearing friction due to the vacuum, and the risk to electronic components posed by radiation, all strongly influenced the design of the little ROKVISS robot. All components which generate heat had to be connected with the casing in a way that allowed heat dissipation. **Latchup protection** (patent pending) prevents damage to



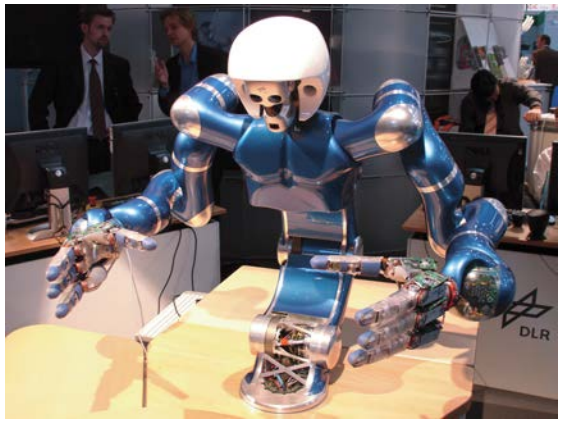


Figure 7.7.13: *DLR's one- and two-armed lightweight robot with multiple-fingered hands is the basis for future robonauts (Source: DLR).*

the electronics from ionizing radiation. This circuit monitors the power drawn from individual electronic components (DSPs, FPGAs and memory modules). When bombarded by protons, the power consumption of these components increases instantaneously. The latchup protection circuit then turns off the power supply to the affected joint electronics and

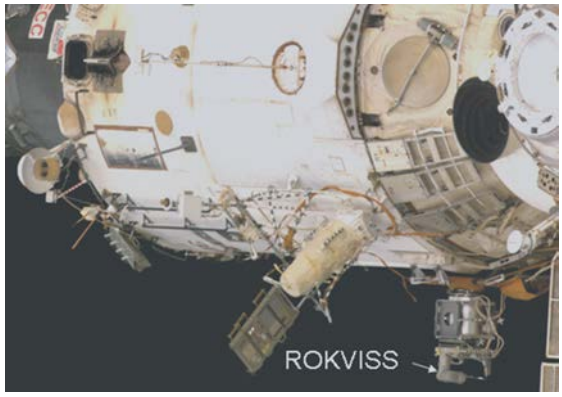


Figure 7.7.14: *ROKVISS on the ISS (Source: DLR).*

at the same time eliminates the charge stored in the capacitances on the circuit board. After an interval of about 2 ms the power supply is reactivated and operations can continue.

During operation of the joints special **monitoring software** supervises the operations software and its parameters. If changes caused by external influences (radiation) are identified, the affected location is automatically repaired. These two protection mechanisms have so far assured problem-free operation of ROKVISS (Figure 7.7.14).

Standard electronic components with an extended temperature range (−45 to +85 °C) were incorporated into the **ROKVISS joint electronics**. These are six to eight times less expensive than space-qualified components. One exception is the latchup protection, which is composed of radiation-tolerant components (Figure 7.7.15).

ROKVISS has its own **S-band communications unit** and is accordingly almost independent of the ISS data transmission systems. With this direct communication link ROKVISS telepresence experiments can be directly controlled when the ISS passes over the DLR ground station in Weilheim. The contact times are up to 8 minutes, depending on the orbit. The robot in space receives its movement commands from an operator on the ground. A force-sensing joystick functions as a **haptic input device** for the operator. The contact forces which arise as the robot interacts

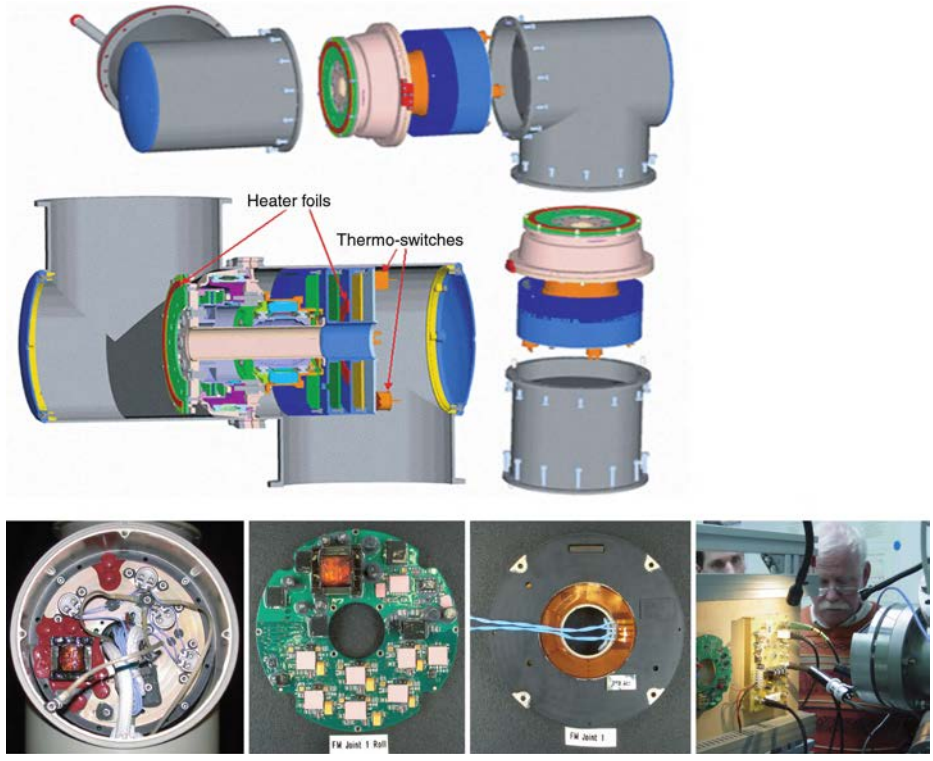


Figure 7.7.15: *ROKVISS with two lightweight joints and innovative latchup electronics (Source: DLR).*

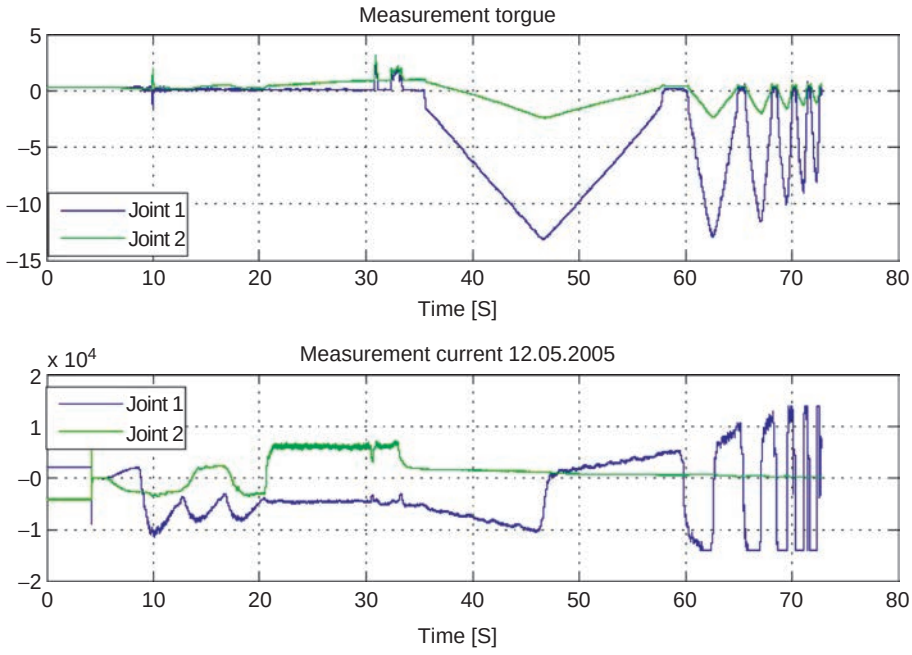


Figure 7.7.16: *Measurement data on the behavior of the joints (Source: DLR).*

with its environment are conveyed to the joystick on the ground. At the same time images from the stereo camera are displayed at the ground station. The operator can observe and sense the interaction; the operator feels telepresent. This telepresence makes extremely high demands on the **data communications path** as to jitter and total signal propagation time. The data clock rate should accordingly be at least 500 hertz and the total signal propagation time, as mentioned above, less than 500 milliseconds. In the meantime, measurements during the ROKVISS experiments have shown that the total signal propagation time is under 20 milliseconds. This is considerably better than originally expected and represents an absolute novelty in space robotics.

The experiment has been in operation since March 2005 and is providing valuable measurement data about the behavior of joints in free space and the efficiency of telepresence operation.

The friction in the joint drive system can be determined by the measured torque, the joint position at the gearbox and the amount of power drawn (Figure 7.7.16). There was an increase in gear friction of about 50% compared to measurements taken immediately before the launch, but no significant change thereafter. This suggests that the properties of the space-qualified lubricant (conventional lubricants would evaporate in a vacuum) changed because of the ambient free space conditions, but then achieved a stable operating state.

The ROKVISS experiment was originally planned for one year. But since the manipulator system continues to perform outstandingly, this very successful mission has been extended a further year. The experience gained from ROKVISS is now being applied to the planning for future space robot missions.

A return to the subject of **direct remote control** is in order here. Because a robot grasping element has six degrees of freedom of movement in space, the **trackball** developed for ROTEX (now called the space ball or space mouse) is very suitable for intuitive human–robot interactions in six degrees of freedom in space. Normally **velocity control** is realized; the small displacements of the control cap are proportionately converted to velocity commands, but force and momentum instructions are also possible using the same concept since the stiffness of the control hardware is known. Joystick displacements are also usually interpreted as velocity commands, but only two degrees of freedom are completely

intuitively controlled in this way. If such a joystick is equipped with two small actuators, then the user can supply a distant slave robot with a force, as in the case of ROKVISS (Figure 7.7.17). But if one wants to control six degrees of freedom with feedback, then a force-reflecting hand controller with at least six actuated joints, actually a small torque controlled robot (master), is necessary. Master and slave can have completely different kinematics. The first space-relevant concepts in this direction were developed at JPL in Pasadena, California, in the 1980s, but have so far not been employed (Figure 7.7.19).

If one wants to remotely control robot arms with artificial hands in an especially intuitive and efficient way, then so-called **exoskeleton techniques** using data gloves are a possibility which is also being investigated by ESA, for example. In this case the entire human arm is involved in the control and force feedback (Figure 7.7.20). This makes it possible intuitively to remotely control free-flying robonauts (Figure 7.7.21); the technique will presumably be employed for future orbital servicing (see also Section 7.7.2).

NASA is also developing two-armed humanoid torsos as a basis for future robonauts (Figure 7.7.22).

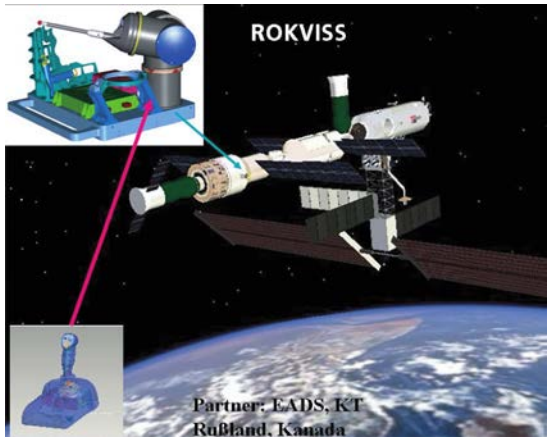


Figure 7.7.17: *ROKVISS on the ISS (Source: DLR).*

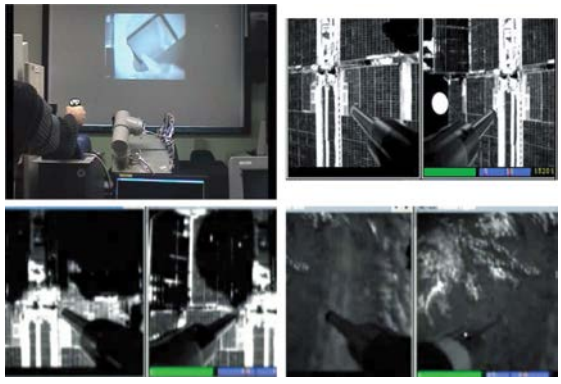


Figure 7.7.18: *Telepresence control (upper left) and the camera images (stereo) of the ROKVISS arm (Source: DLR).*

7.7.1.4 Lander Missions for Exploring Space

The successes experienced with unmanned missions to the planets and other celestial bodies during the last 25 years have been remarkable and have become absolute highlights of space flight. Special challenges



Figure 7.7.19: *The first force-reflecting hand controller at JPL (Source: JPL).*

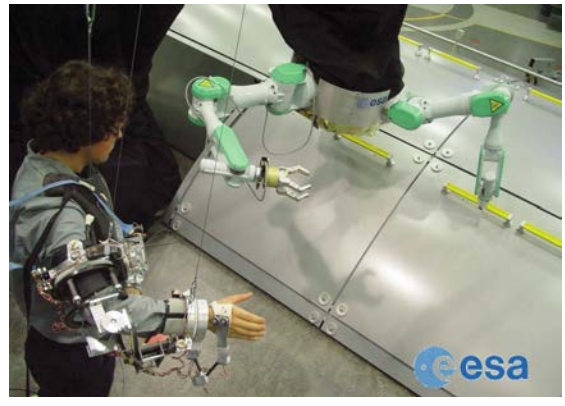


Figure 7.7.20: *Exoskeleton developed by ESA (Source: ESA).*

are of course posed not only when probes fly by or orbit celestial bodies (the so-called **orbiters**), but if a so-called **lander** (actually, an entire landing system) separates and descends to a surface, particularly if after landing a mobile system (on wheels, and usually called a **rover**) sets out to investigate its environment. The first systems of this type worldwide were the Russian lunar spacecraft Lunokhod 1 (1970) and Lunokhod 2 (1973) (Figure 7.7.23).

The last one traveled an impressive 37 km on the lunar surface. The eight-wheel system (with a ninth wheel to measure movement, an odometer) was remotely controlled via **teleoperation** using stereo images. Elementary commands such as "turn," "forward," "stop," and "backward," as well as several ground operators representing various hierarchic levels (such as driver, navigator, orbit planner, commander, etc.), were used.

This kind of remote control is not possible on **Mars** because of the long signal propagation time of 15 minutes. In the meantime, NASA's JPL in Pasadena has gathered considerable experience with semiautonomous planetary rovers, especially the **Mars rovers**. The Sojourner rover (Figure 7.7.24), a key element of the 1997 Mars Pathfinder mission, was the first autonomous vehicle to travel over the surface of a planet. It used the stereo camera on the landing mast, whose

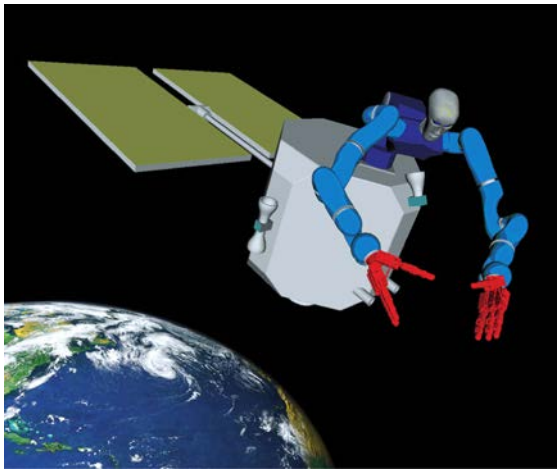


Figure 7.7.21: *DLR conception of a "free-flying robonaut" (Source: DLR).*

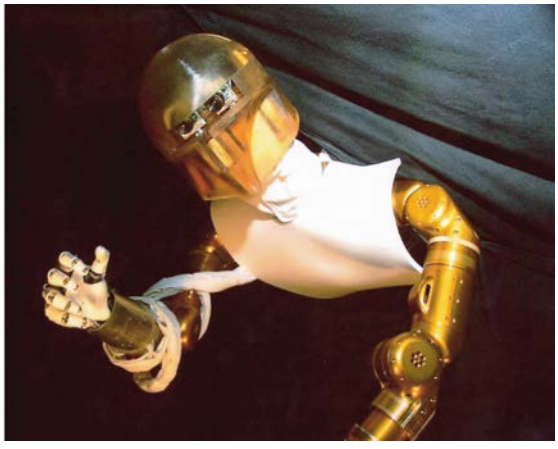
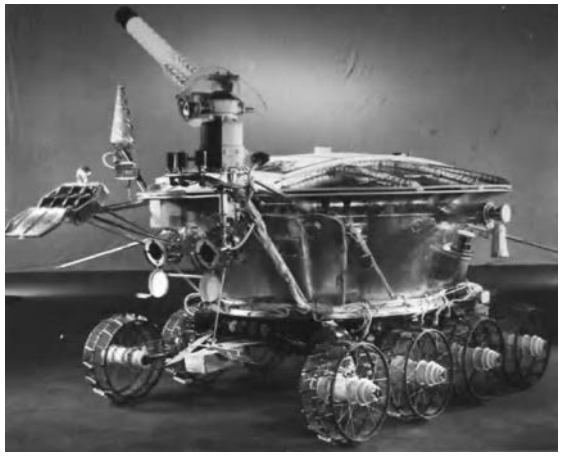


Figure 7.7.22: *The NASA robonaut (Source: NASA).*



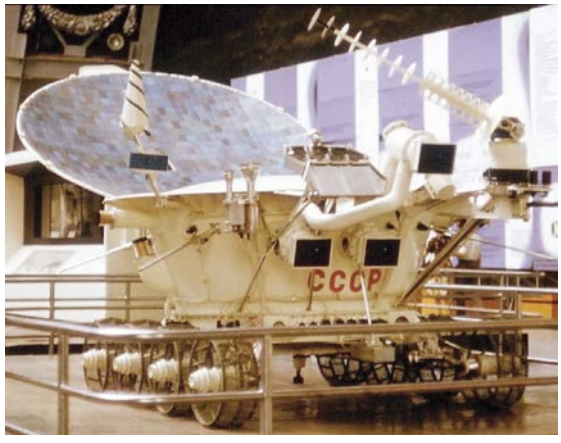


Figure 7.7.23: *Lunar vehicles Lunokhod 1 (above) and Lunokhod 2 (below) (Source: NASA, [7.7.6]).*

images facilitated the planning of approximate route milestones by the ground operators. Sojourner had a risk-detection and collision-avoidance system to guide it safely through a stony field from one specified target to the next. It investigated a dozen distinctive rock fragments using customized spectrometers (Figure 7.7.25).

The Sojourner success promptly led to a decision to conduct a follow-up mission, MER (Mars Exploration Rovers), with the two rovers *Spirit* and *Opportunity* (Figure 7.7.26).

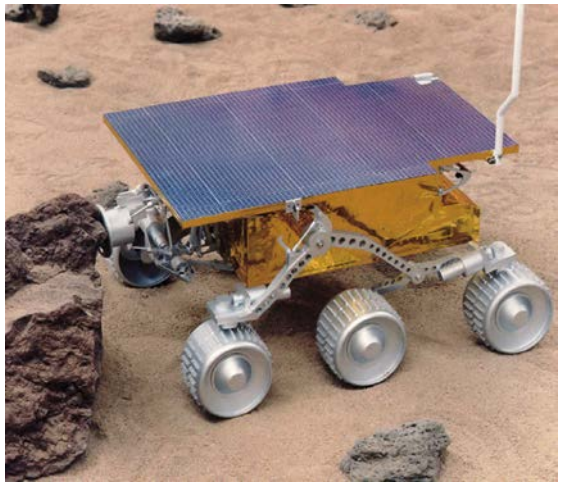


Figure 7.7.24: *The Mars rover Sojourner (Source: NASA).*

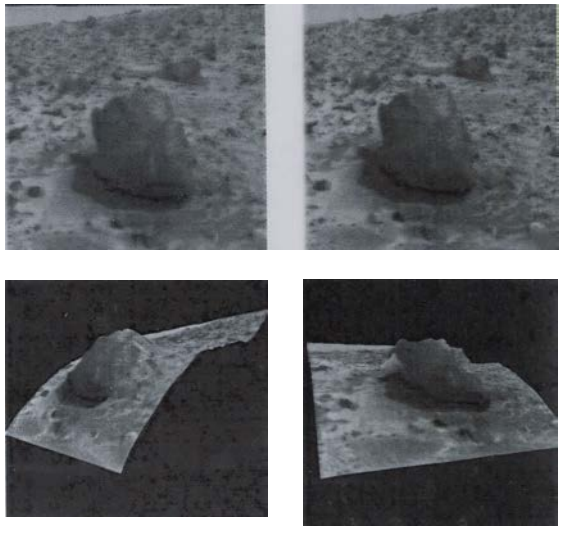


Figure 7.7.25: *Neural nets generate a 3D model of a YOGI stone from the stereo images (Source: JPL).*

These two Mars rovers have been traveling on the surface of Mars since 2004, far longer than originally planned. They are the first Mars rovers with a robot arm (five degrees of freedom) which can move instruments and cameras. These arms use contact sensors to avoid destructive influences caused by environmental forces. Similar to the Sojourner mission, route targets are generated on the ground but are no longer entirely dependent on a stationary lander camera system.

New Mars concepts developed in Russia in recent years have not yet been implemented (Figure 7.7.27).

The great success of the MER mission will certainly spur the dispatch of **sample–return rovers** to Mars. Such rovers collect rock samples so that they can later be brought to Earth, and this kind of mission has been planned for several years.



Figure 7.7.26: *The Mars rover Spirit (Source: JPL).*

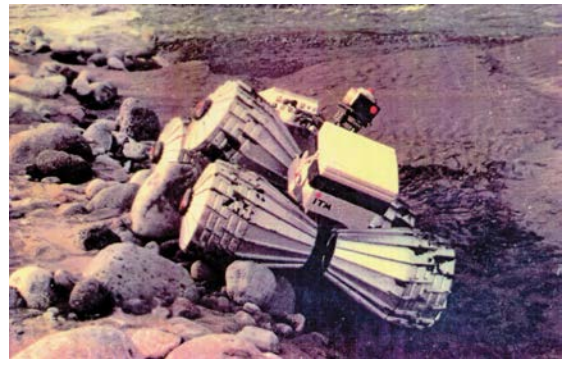


Figure 7.7.27: *Russian rover concept with six conical wheels (Source: VINI Tansmash).*

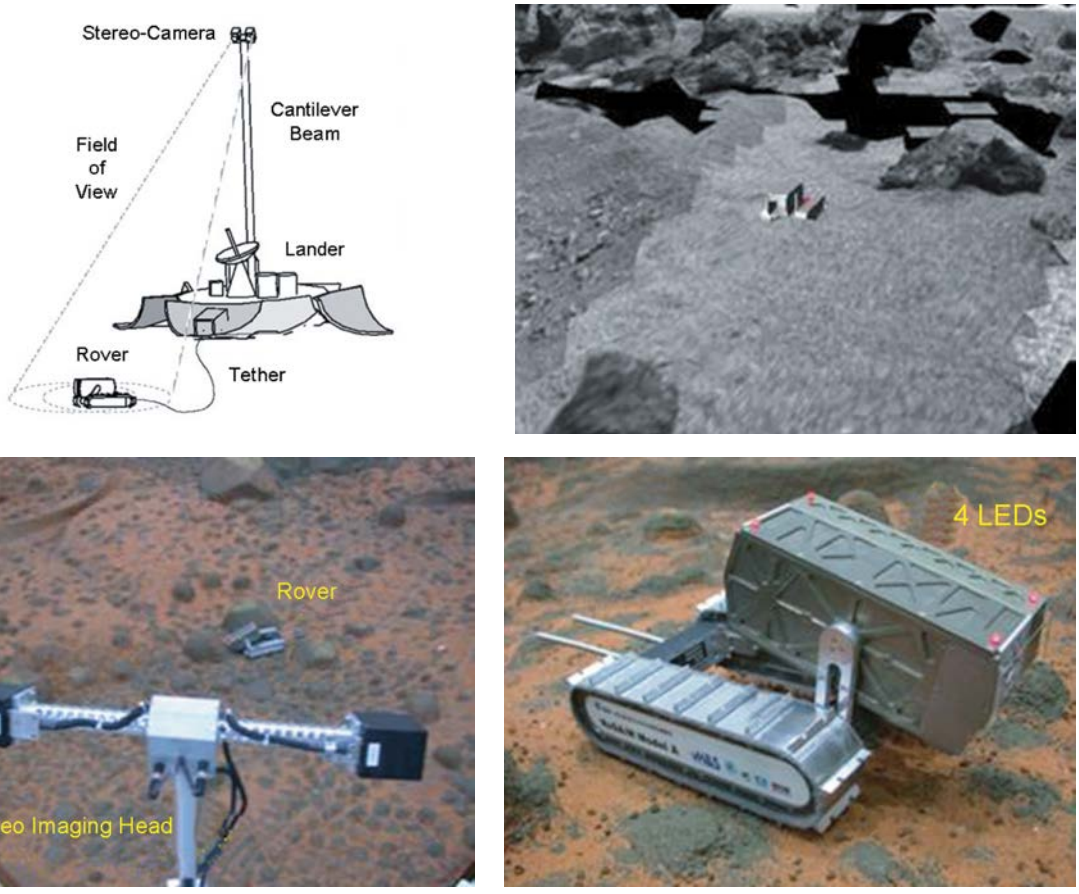


Figure 7.7.28: *A rover navigation system developed for ESA by DLR which incorporates a stereo camera on the lander (Source: DLR).*



Figure 7.7.29: *DLR's crawler: simulation and reality (Source: DLR).*

Compared to NASA projects, the plans of ESA are relatively modest. The ExoMars mission scheduled for after 2011 will involve a rover which may resemble the MER rovers designed by NASA. DLR has developed a lander-based navigation system with tiny LEDs on the rover as part of an ESA study (Figure 7.7.28).

As an alternative to wheeled rovers, **walking machines** like the "crawler" (Figure 7.7.29) developed at DLR may be of interest. It is even possible that a combination of wheels and legs might turn out to be the optimal configuration because of the versatility offered.

The field of robotics can also demonstrate great progress in developing **algorithms for 3D world modeling** from stereo images. They are even suitable for 3D surface modeling with an orbiter camera like the HRSC stereo line camera of DLR, which has been orbiting Mars for the past few years as part of the Mars Express mission and is modeling that planet for the first time three-dimensionally at a spatial resolution of 10 to 20 m (Figure 7.7.30).

Robotic experts make no secret of the fact that their **future vision of Mars exploration** resembles what is depicted in Figure 7.7.31, with mobile robots which can collect rock samples and analyze them on location, and perhaps even transmit real-time stereo images to Earth. Instead of only a few astronauts, countless people in stereo projection rooms on the ground could feel as if they were "telepresent" on Mars.

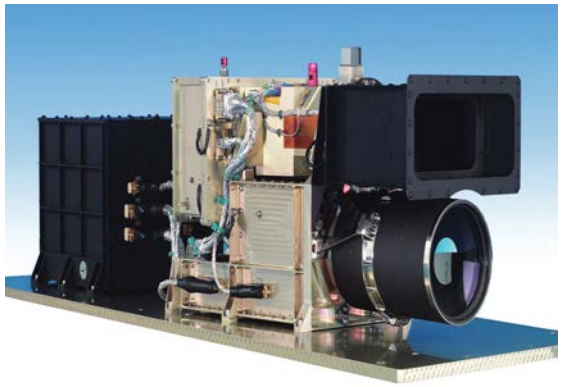




Figure 7.7.30: *Ice-Craker and Mars with Mars Express Camera (Source: DLR).*

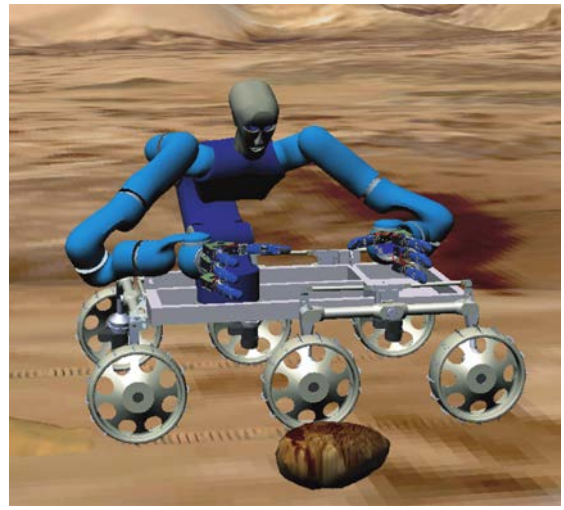


Figure 7.7.31: *Robotic torso on a Mars vehicle (Source: DLR).*



Figure 7.7.32: *JPL is studying an autonomous planetary airship (Source: JPL).*

It should also be mentioned that responsible institutions like JPL are also considering using **slow-flying platforms**, namely autonomous airships (Figure 7.7.32), on objects like Saturn's moon Titan. And, finally, **lunar missions** have recently returned to the focus of attention. The establishment and operation of permanent stations there will assuredly not happen without robotic support.

7.7.2 On-Orbit Servicing

On-orbit servicing (OOS), the provision of services in space, is a key element in the continued exploitation of the space and in establishing and maintaining the required space infrastructure. It can considerably reduce the operating costs for contemporary unmanned space assets such as navigation satellites and geostationary communications satellites. This aspect is already an important consideration in human space flight and is usually carried out by the astronauts themselves – sometimes supported by hand-controlled robotic elements. For unmanned space activities it means the use of robots with a high degree of autonomy. This aspect of unmanned space flight, which is becoming increasingly important and necessary, is discussed in this section.

The **servicing** of satellites in orbit includes all aspects of component assembly and equipment maintenance (both preventive and corrective), the replenishment of consumables, and upgrade and repair. The life cycle of a space mission can be subdivided into three segments, each with its own typical sources of failure:

1. If the payload cannot be injected into the specified target or transfer orbit because of failure in one of the launcher stages, then orbit correction may be required. In most cases the satellite cannot accomplish this on its own, therefore an orbit transfer vehicle could provide support.
2. The "mission start" phase includes at least the in-orbit checkout, but can also extend 6 to 12 months into the operational phase. A failure message during so-called begin of life (BOL) can imply the necessity for repair, such as the subsequent release of a jammed deployment mechanism for a solar generator or an antenna dish.
3. The "mission conclusion" phase can – but must not – be identical to the end of planned service life. A satellite can reach its so-called end of life (EOL) prematurely because of wear or equipment obsolescence.

A typical **servicing profile** is usually associated with BOL failures, which may affect the entire mission. In the case of EOL failures (aging), the main interest is in increasing the lifetime beyond EOL; for example, in order to maintain the services provided by the satellite. Scheduled servicing operations require a servicing strategy between BOL and EOL, such as module or component upgrade, replacement or retrofit or the refurbishment of functional surfaces like reflectors, solar generators, radiators, etc. With the help of the Space Shuttle, astronauts carried out the only major servicing program to date for an unmanned satellite at considerable cost: the repair and servicing missions for the Hubble Space Telescope.

7.7.2.1 Characteristics and Options

Servicing and repair by robotic elements in space involves certain basic task requirements: access to the orbit of the target object, orbital maneuvering, and proximity servicing operations.

Servicing is chiefly carried out by the replacement of modular components for purposes of repair, maintenance or upgrading. The accommodation of all replaceable components in a single replaceable module minimizes both the complexity and cost of servicing. **Modularizing** has another advantage: the complexity is directly proportional to the number of active interfaces between system components. One can group typical service missions into three main categories, which can be further broken down into subcategories, and then derive the requisite robotic capabilities for each field of operation:

1. **Inspection:** Recording and collecting data from the target satellite. This task requires merely the capability to fly around the satellite plus the corresponding visual sensor technology:
   (a) **Remote Inspection:** Failure diagnosis by remote sensing of target.
   (b) **Close-up Inspection:** Scanning the target satellite using various kinds of sensor technology.

2. **Motion:** Translation services to the target satellite, which requires docking capability. These tasks

relate to the assumption of orbit and attitude control, or orbit and station change. The service satellite must accordingly have an appropriate propulsion and attitude regulating capability:

(a) **Reorbit:** This task includes the transfer of a stranded target satellite to its originally planned operational orbit, or a later change of operational orbit in case the former position is to be occupied by a new or superior satellite.
(b) **Deorbit:** This task includes the transfer of the target satellite into a so-called graveyard orbit, or into a reentry orbit for its deliberate destruction and thus disposal in the atmosphere.
(c) **Salvage:** This task involves either the transfer of the target satellite to another space vehicle, or its nondestructive reentry.

3. **Manipulation:** This involves dedicated control of subsystems or some type of intervention, which requires additional interfaces between the target and service satellites:
   (a) **Maintenance and Checkout:** Supplying the target satellite with consumables (liquids, fuel) as well as cleaning, resurfacing and decontamination tasks.
   (b) **Repair:** Diagnosis and correction of module failure in the solar generators, gyroscopes, antennas, etc.
   (c) **Retrofit:** Upgrading by replacement with more efficient modules.
   (d) **Docked Inspection:** Failure diagnosis by physical interrogation of the target satellite via connectors.

All these tasks can be necessary throughout the entire lifetime of the satellite (see Table 7.7.1). Exceptions are maintenance and retrofit in the BOL phase, and reorbiting and repair in the EOL phase.

Table 7.7.1: *Overview of servicing tasks.*

Type of service	Demand	
	Emergency	Scheduled
Remote inspection	●	●
Docked inspection	●	●
Reorbiting	●	
Deorbiting	●	●
Salvage	●	●
Maintenance		●
Repair	●	
Retrofit		●

7.7.2.2 On-Orbit Services

There is a **natural hierarchy of tasks** which can be carried out by on-orbit servicers in order of increasing complexity. **Inspection** involves investigating space assets for damage assessment and failure reporting of the externally visible state of the target. This requires the technology of **highly precise maneuvering**. For greater distances in the kilometer range, absolute navigation may be necessary, but in all other cases the preference, for safety reasons, is precise maneuvering with relative navigation to the target, whereby the attitude and orientation of the service satellite have to be determined with appropriate sensory instruments.

In order to carry out the next phases in on-orbit servicing, a **rendezvous and docking** maneuver is necessary. In addition to the previously mentioned navigation capability required for remote inspection, a very precise resolution in determining the relative position of the two vehicles is particularly essential for docking maneuvers. This can be accomplished with exact distance measurements, but, ideally, with the aid of optical cameras and appropriate image processing. Whether a docking procedure is handled automatically or controlled from Earth mainly depends on the communication links from and to the service satellite, which are determined by the orbit.

Target satellites can be of constant attitude and status if their attitude control system is functional. This significantly reduces the effort involved in tracking the target and especially in docking maneuvers. A deactivated satellite may have an axial rotation around its longitudinal axis (a typical state after the so-called passivation) or may be tumbling randomly. In either case high demands are made on the maneuvering, navigation and docking capabilities of the service satellite. As a rule, such objects are approached only

in emergency situations, in order to retrieve them or to initiate controlled atmospheric reentry.

Docked inspection is the basis for accessing the inside of the spacecraft to diagnose its condition and failure. So far, the target satellites have usually been passive and unequipped with interfaces for satellite servicing. Investigations are in progress to equip unmanned space vehicles with active interfaces in order to establish a variety of connections between target and service satellite for carrying out servicing operations.

The next level of complexity in on-orbit servicing is **moving the target satellite** with the help of the docked service satellite, usually to take over the task of attitude and orbit control, or to deliberately change its orbit. The precondition is a successful docking procedure. In order to move the combined system, the service satellite undertakes variable **thrust vectoring** in order to efficiently utilize the available fuel, taking into account the momentary location of the center of mass of the combined system. The amount of propulsion power is not decisive; rather this influences the inertia of the attitude control system.

The most profitable application for the already mentioned and yet to be mentioned types of on-orbit servicing is docking with telecommunications satellites at EOL and taking over their attitude and orbit control for a number of years in order to maintain profitable telecommunication services. Secondary to this is the salvaging of new, stranded telecommunications satellites which, because of the failure of a launch vehicle stage or of their own apogee motor, cannot reach their target orbit on their own. Such a rescue maneuver only makes sense if these usually expensive satellites are put into their target orbit at a cost that is considerably lower than the otherwise expected insurance sum.

Whereas technical solutions for the systems named so far (also for overall systems) have already been devised, some of which have even been demonstrated, the following on-orbit services are yet to be realized since they require appropriate **active interfaces**. As mentioned, only the next generation of satellites is apt to have such interfaces. In any event, the idea is to have as many types of interfaces as possible, such as electrical, mechanical, thermal, radiofrequency and also interfaces for transferring fluids and gases. In particular, **refueling** of orbiting spacecraft is probably the most lucrative aspect of the named scenario because it fully restores the target satellite function and, after the refueling procedure, the service satellite is free to attend to other tasks at other targets. The automatic handling of connectors in orbit is technically possible if the respective robotic capability is fully available. The capabilities of robots on the ground are well known and have been demonstrated. But to carry out these kinds of tasks in orbit, the technologies have to be adapted to different marginal and environmental conditions. Reliability must also be extremely high since the service satellite should not itself become a repair case.

Besides the automatic handling of connectors, the exchange of modules is probably the last link in the chain of applications described here. The service life of a satellite is defined not only by its fuel supply and consumption, but also by the lifetime of its subsystems. If such subsystems can be so integrated into satellites in the future that they are accessible from outside, they could be exchanged by a robot arm. Examples are reaction wheels and gyroscopes which are subject to mechanical wear, or even the periodic replacement of precision clocks on-board the Galileo satellites to keep them at the state of the art and to increase signal precision.

In the context of on-orbit servicing, the automated assembly of large structures and instruments which cannot be launched in one piece or put into orbit and then unfolded should be mentioned. This future scenario of automation in space with the help of robotics will in the long run determine accessibility to outer space.

The first consideration when designing service satellites is to decide between **single service** and **multiple service capability**. The constructions differ in architecture, logistic support and financing models. There is little doubt that multimission service satellites are more useful for emergency operations. Many configurations include in-orbit depots for consumables or supplies. From the servicing perspective, the ISS offers an ideal platform from which to initiate on-orbit servicing or provide an ideal environment for the servicing itself, as it

Table 7.7.2: *Classification of the various types of on-orbit services.*

Class	Observation	Movement	Manipulation
Service	• Remote inspection	• Reorbit maneuvers (station keeping) • Deorbit • Salvage	• Maintenance • Upgrading, repair • Docked inspection
Purpose	• Status information	• Prolongation of operations • Availability of orbital slot • Avoidance of uncontrolled deorbiting • Reuse	• Expanded operations • Improved function and efficiency • Information
Value	• Low • Low cost for multiple clients	• Medium	• High
Incidence	• High	• Medium	• Low

makes possible a combination of automation and astronaut support (Table 7.7.2).

7.7.2.3 Components

A complete service system constellation with a complex mission has many aspects, which can be described as follows:

- **Mission Configuration:** A suitable operations scenario has to be identified before the other elements of a service mission can be defined.
- **Service Satellite Platform:** The design of this spacecraft depends on the operations scenario. For every rendezvous and docking mission it is probable that manipulators will connect the spacecraft to the target satellite. In addition, navigational capability and a certain degree of autonomy are desirable for most operating scenarios,
- **Service Satellite Payload:** This payload is strongly dependent on the particular mission purpose and can have such elements as docking mechanisms, sensors, actuators, manipulators, control software, etc.
- **Target Satellite:** This satellite usually determines the mission and operations scenario. The first missions should be based on already existing satellites. Only in the future target satellites will be equipped with on-orbit servicing elements.
- **Launch:** As a satellite launch is an extremely expensive affair, an important element in the configuration of operations scenarios is cost reduction. Costs can be reduced by using a single (or reusable) low-cost launcher. Another possibility to reduce costs would be to use dual-launch systems or a piggyback solution, as with the Ariane 5 rocket,
- **Ground Segment:** On-orbit servicing requires a suitably equipped ground station or a ground station network to monitor and control the service satellite.
- **Logistics:** Finally, the global on-orbit servicing scenario determines the number of service satellites required in orbit and their parking position, as well as what supplies or replacement modules are to be stored and carried.

7.7.2.4 Technological Requirements

The technologies required for the on-orbit components described above must meet certain requirements:

- **Payload:** The robotic elements required for each particular mission scenario must satisfy high demands as to precision and reliability. Such systems have long been available in the laboratory. The challenge is to make the complex robotic system suitable for the space environment. In most cases there is no redundancy for primary elements like the docking mechanism and manipulator arm. In order to increase reliability, elaborate ground testing under the most realistic conditions possible as well as the use of robust components are therefore necessary. The specialized fields to be addressed are electronics, mechanics and mechatronics,
- **Sensor Technology:** In order to carry out in space the operations described above, corresponding sensor equipment has to be on-board. It serves on the one hand to grasp the target satellite in the rendezvous phase and on the other hand to determine as precisely as possible after approach the attitudes and relative positions of both the service and target

satellites during docking or close-up inspection. It is therefore desirable to work with optical cameras which can be used for object recognition as well as inspection tasks.

- **Platform:** There are high requirements placed on the Δv budget, which directly determines the lifetime and efficiency of the service satellite. The platform also requires an extremely precise and adjustable propulsion system when near the target and for the docking maneuver. In addition, all the classic elements have to be on hand to control the service satellite and maintain communication with the ground station. If video data is used, an appropriately powerful ground link has to be available.
- **Software:** High demands are placed on the software algorithms which must be able to close the control circuit between the sensor and actuator technology (attitude regulation, mechanisms), for some operations, fully autonomously and in real time. Rapid processing of sensor data and comparison with available models are key elements in autonomous robotic activities in space.
- **Operations:** Anything which the service satellite cannot carry out fully autonomously must be controlled by appropriate operation commands from the ground. This includes primarily the complete approach and rendezvous phase as well as the first part of the docking phase. Since there is always a high risk of collision, the design and verification of these operations must be carried out with extraordinary care.

7.7.2.5 Influencing Factors

The organization of a service mission is determined by many important factors:

- **Requirements on the Service Satellite:** These depend on the particular servicing scenario. For a nongeostationary constellation in several orbital planes (e.g. Galileo), fuel limitations require at least one service satellite per orbit plane. But the number of target satellites and the complexity of the servicing tasks could make it necessary to have several service satellites in each plane.
- **Frequency:** This depends on the planning of nominal events and assessment of unscheduled events. If maintenance is an integral part of the mission (attitude regulation, refueling, regular upgrades), frequent visits are essential for the success of the mission.
- **Rendezvous and Docking:** RvD places high demands on the design and the payload. The number and type of robotic elements have to be defined. In addition, the docking speed influences the design of the docking mechanism. Low speeds are desirable to avoid damage, but they require more fuel. An appropriate sensor regime is also needed to improve the precision of the docking maneuver and thereby reduce the probability of collision.
- **Orbit and Inclination:** The design is primarily determined by the particular orbit. Service satellites crossing the Van Allen belt, in some cases several times, require additional radiation protection, which affects cost and mass. The number of launch opportunities is also an important consideration since dual or piggyback starts are less expensive. There are frequent launches particularly for the geostationary transfer orbits, and many satellites in geostationary orbit are potential on-orbit servicing candidates.
- **Availability:** Servicing according to a schedule is easier, since the service satellite only has to transport equipment and fittings for defined situations and it is known in advance which operations are to be performed. An emergency operation is, by comparison, more difficult since it could be necessary to first carry out inspections and failure search, followed by repair as necessary and possible, tasks for which the client satellite may not be designed (e.g., cutting through insulation layers).
- **Logistics (Consumables):** Logistics are an important factor in a servicing mission. For example, a servicer can be constructed to transport one or two replacement modules or even large, completely full fuel tanks for refueling operations.

7.7.2.6 Past and Planned Mission Examples

A brief survey follows of past and future missions which include servicing tasks in space. Although completeness is not claimed, the survey is, however, fairly representative.

SNAP-1

One of the first European service satellites, which only had the task of inspection, was the British SNAP-1.

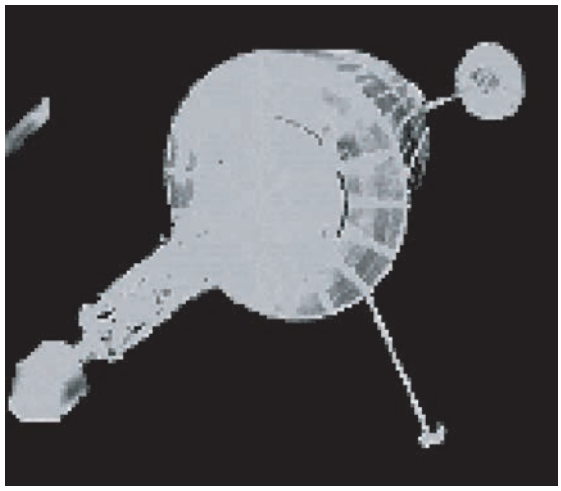


Figure 7.7.33: *Snapshot of the Russian Nadeshda satellite, made by SNAP-1 (Source: Surrey Space Centre).*

This **nanosatellite**, designed and constructed by Surrey Satellite Technology Ltd, was a very small, but because of its great agility, fully functional **inspection satellite** which used a camera to inspect an object from several sides. SNAP-1 was launched in 2000 and made successful close-up approaches of the Russian Nadeshda satellite (Figure 7.7.33) as well as, from a considerably greater distance, of the Chinese Tsinghua-1 satellite. This mission also served to demonstrate nanosatellite technology and had a relatively small budget.

Such free-flying **inspection systems** were considered especially for the ISS, which would have been a very interesting and certainly useful application. But to date none of the plans have been realized, both because of lacking financial support and, more importantly, because the risk that such a system might collide with the station could not be fully eliminated, and the safety of the station and its crew has priority.

DART

NASA's so-called Demonstrator for Autonomous Rendezvous Technology (DART) was an experiment to demonstrate new technologies in orbit necessary for **localizing and approaching a target satellite** (Figure 7.7.34). Up to that point such tasks had been carried out exclusively by astronauts as part of human missions. The declared purpose of DART was, for the first time, to have these tasks carried out completely autonomously by an unmanned satellite. The target satellite was a retired US communications satellite named MUBLCOM. The mission was launched in 2005 and was, unfortunately, unsuccessful. Because of problems with the fuel supply there was an unnoticed and uncontrollable collision with the target satellite. The main reason for the failure of the rendezvous phase was **navigation based on GPS**, which turned out to be unreliable for such a mission task.



Figure 7.7.34: *DART (Source: NASA).*

ETS-VII

The Japanese ETS-VII mission (Engineering Test Satellite; Figure 7.7.35) of 1997 is to date the most complex successful **technological demonstration of a service mission** [7.7.1]. It was carried out in a 550 km altitude orbit under the responsibility of the Japanese space agency NASDA (now JAXA). The service satellite, named ORIHIME, weighed 2.5 tonnes and the target satellite, HIKOBOSHI, 0.4 tonnes. The manipulator arm was 2 meters long and had six degrees of freedom. With the help of various available robotic technologies, basic tasks were tested like autonomous rendezvous and docking with a target satellite, control of the manipulator arm via teleoperation, and force-feedback control. Germany has contributed its GETEX experiment to the ETS-VII mission, see Section 7.7.1.2.

Orbital Express

Orbital Express (Figure 7.7.36) was a project financed by DARPA (Defense Advanced Research Projects



Figure 7.7.35: *ETS-VII (Source: NASDA).*

Agency) to demonstrate **refueling** of a military satellite in space as well as the **replacement of certain components** and the realization of **repairs** using a manipulator arm after a successful docking maneuver [7.7.2]. The mission was launched in March 2007. Astro as the service satellite and NextSat as the target satellite were brought into orbit at the same time and then separated in order to initiate the actual technology demonstration, beginning with rendezvous and docking. This elaborately prepared project represents a complete on-orbit-servicing scenario and is thus a milestone in the history of space activities. Especially, the **joining of different types of connections** (mechanical, thermal, electrical, radiofrequency, liquid, gaseous) posed enormous challenges, in addition to the complex tasks of rendezvous and docking and controlling the manipulator arm.

In Europe an ESA study is currently underway to investigate criteria involved in joining the different connections of a service and a target satellite, but in this case with the sole goal of defining future interfaces and identifying appropriate technical concepts for next-generation communications satellites.

FREND (SUMO)

FREND, formerly SUMO, is the designation for the second DARPA service mission besides Orbital Express. Its primary goal is to demonstrate the **maneuvering and repositioning** of orbiting satellites [7.7.3]. Various aspects of service life prolongation and controlled deorbiting are also to be investigated. A significant feature is that the service satellite uses manipulator

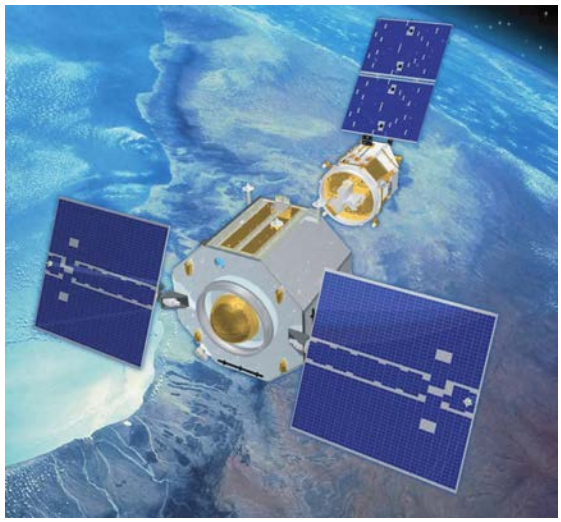


Figure 7.7.36: *Orbital Express (Source: DARPA/Boeing).*

arms to capture the target satellite without requiring interfaces to the target satellite, so that any target satellite can be approached. This mission accordingly represents another improvement, especially when compared to Orbital Express, which made use of docking interfaces on the service and target satellites. FREND also has the goal of carrying out such operations for the first time in geostationary orbit.

TECSAS and DEOS

Together with foreign partners, in 2006 Germany carried out a first study for a robot-supported service mission named TECSAS to demonstrate technology in orbit (Figure 7.7.37). This mission was similar to the Orbital Express mission launched in March 2007. It consisted of a service satellite and a target in orbit to be connected with the help of a robot arm.

For programmatic reasons this project was not continued, but it is being redefined under the name DEOS and will be a purely national mission. In contrast to TECSAS the robot arm for manipulation and capturing of an uncontrolled satellite will be supplemented by a **docking tool** developed by DLR which links the service satellite and the target in space. As was originally planned for TECSAS, DEOS is to operate in LEO, which makes special demands on the communication link with the ground because only short contact times exist during each



Figure 7.7.37: *TECSAS (Source: DLR).*

pass. The possibility of carrying out communications via so-called relay satellites in order to maintain permanent contact is being considered. This German technology mission will provide many possibilities to test and verify a variety of robotic operations in space and demonstrate the capabilities of German industry and science.

OLEV

A special case of orbital servicing, the service life prolongation of geostationary satellites, is being currently investigated by a European commercial consortium under the leadership of Orbital Satellite Services Ltd [7.7.4]. With the participation of the industrial partners Kayser-Threde GmbH (responsible for the docking payload [7.7.5]), Swedish Space Corporation (responsible for the platform) and the Spanish Sener Ltda (responsible for the hardware and software of the guidance and navigation control system) a proposal with the name of OLEV (Orbital Life Extension Vehicle) is being finalized and is close to realization. The concept was originally developed on the basis of the ConeXpress platform (see Figure 7.7.38) but now a modified platform resembling SMART-1 is preferred, primarily to minimize development costs (see Figure 7.7.39).

This application is based on the capture of operating geostationary communications satellites using

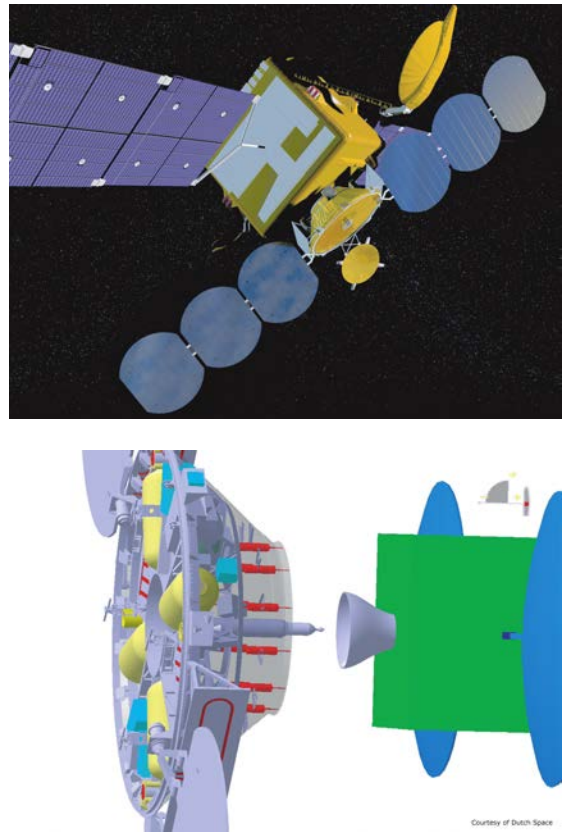


Figure 7.7.38: *CX-OLEV (Source: Dutch Space/Orbital Recovery Group).*



Figure 7.7.39: *SMART-OLEV (Source: Swedish Space Corporation).*

a DLR multisensory capture tool (see Figure 7.7.9) with which practically any of these satellites whose cold gas supply is running low can be grasped via

their always-present apogee nozzle. Their attitude can then be stabilized for years with a set of electric propulsion engines provided by the service satellite. The challenge is not only in the docking and connecting procedure but also, and primarily, in the operations scenario. For a system consisting of two bodies with a common center of mass outside the service satellite, the orbit and attitude regulation has to be carried out using only the engines of the service satellite, which induce unavoidable coupling effects (and the associated losses) for each correction maneuver. These have to be skillfully compensated over an entire orbit. The main difference to the missions described above in LEO is the significantly longer communication period with the service satellite. This makes it possible to command special operational tasks as well as routine control and safety tasks from the ground.

In contrast to the prior examples for technology demonstration missions, the project is entirely commercial. This situation will assure most of the financing through private investors with simultaneous systems design for operational and commercial applications together with one of the major communications satellite operators. Besides service life prolongation, customers also have the option of fleet management to constantly achieve the optimal position over Earth through customized positioning of the fleet vehicles in orbit. Current plans assume a first docking maneuver using a fully operational system in space by 2012.

Bibliography

[7.7.1] Oda, M., Inaba, N. Demonstration Mission of a Satellite Servicing System. Proceedings of the 5th International Symposium on Artificial Intelligent Robotics & Automation in Space, *ESA SP-440*, pp. 221–216.

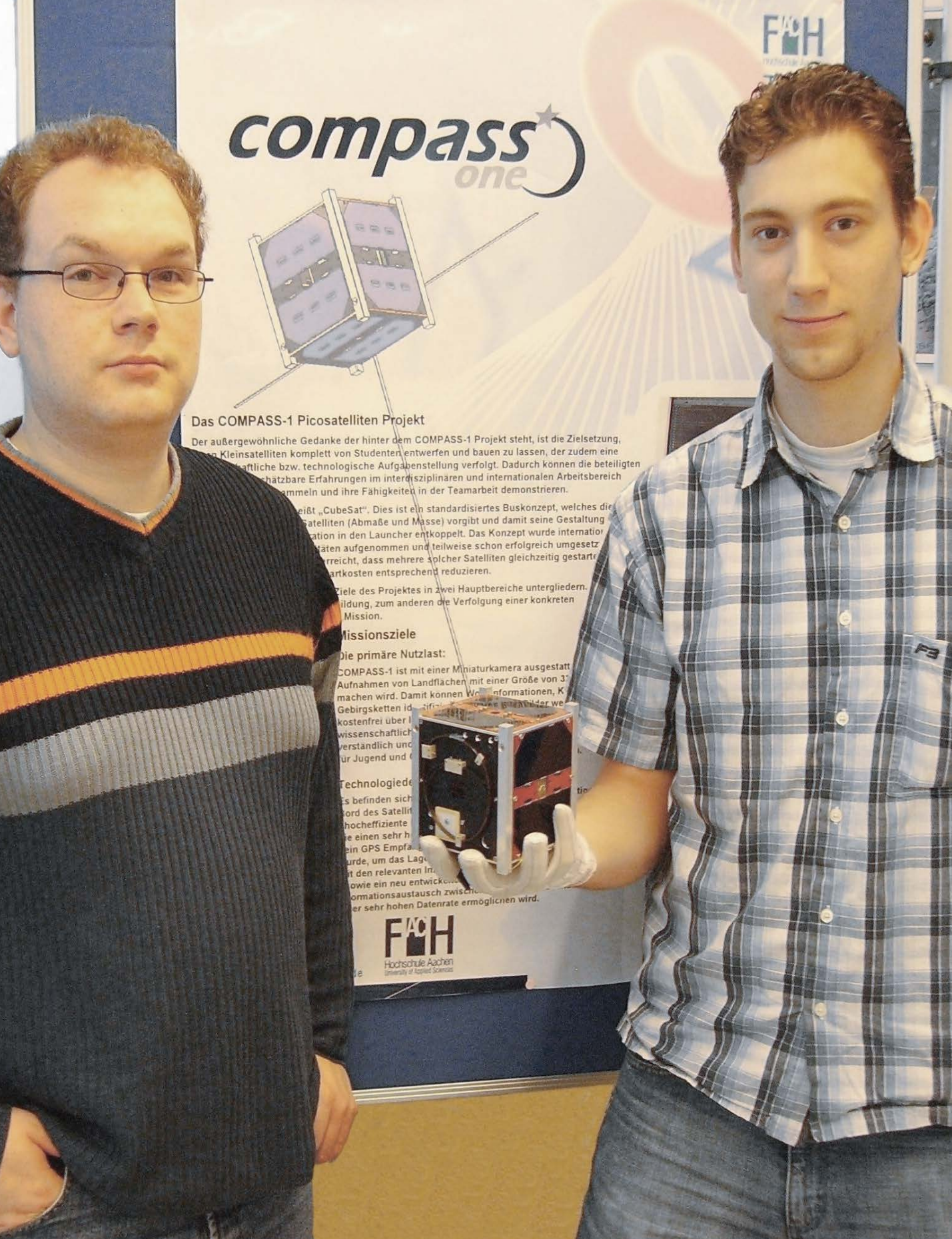
[7.7.2] Satellites Hopes Ride on Orbital Express. *Aerosp. Am.*, February 2007, p. 30.

[7.7.3] SUMO Wrestles Satellites into New Orbits. *Aerosp. Am.*, February 2006, p. 26.

[7.7.4] Kaiser, C., Hofmann, P., Eilertsen, B. et al. On-Orbit Servicing for the Commercial Utilization of Space/Orbital Satellite Services System. Proceedings of the 57th IAF Congress, Valencia, Spain, October 2–6, 2006.

[7.7.5] Kaiser, C., Hofmann, P., Landzettel, K. *et al.* The Docking Payload of the ORS Orbital Life Extension Vehicle. Proceedings of the 9th ESA Workshop on Advanced Space Technologies for Robotics and Automation, ESTEC, Noordwijk, the Netherlands, November 28–30, 2006.

[7.7.6] Horn, J., Reball, S. (eds.) *Moderne Technik von A bis Z.* Leipzig/Cologne: Fachbuchverlag/Verlag TÜV Rheinland, 1991.


compass one
Das COMPASS-1 Picosatelliten Projekt
Der außergewöhnliche Gedanke der hinter dem COMPASS-1 Projekt steht, ist die Zielsetzung,
Missionsziele
Die primäre Nutzlast:
Technologiede
Hochschule Aachen

8 Spacecraft Design Process

Klaus Brieß

8.1 Mission Concept and Architecture

Klaus Brieß

8.1.1 Elements of a Space Mission

For the development of a concept and architecture for a space mission, quite different systems, elements and their interactions have to be planned. Each space mission can be divided into **elements of the space mission**. These are the main components of a space mission. The mission elements vary slightly depending on the character of the space mission and they are sometimes defined differently by the different space organizations and establishments. But most of the elements are defined similarly, as depicted in Figure 8.1.1.

The arrangement of the mission elements and their interactions are characterized by the **architecture of the space mission**. The main elements of a space mission are explained in the following sections.



Figure 8.1.1: *Elements of a satellite mission.*

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9

8.1.1.1 The Mission Objective

The mission objective includes the reason for and purpose of the space mission, the goal and the subject of the mission, as well as its benefits. The mission objective can be, for example, operational weather monitoring for weather service organizations or investigation of phenomena on the Earth's surface or in deep space for scientific establishments or other institutions. In these cases the mission subject would be the Earth's surface or deep space.

8.1.1.2 The Mission Concept

The mission concept includes the development, integration and testing, implementation, operations and the disposal of the space mission. It comprises the following:

- Concept for the accomplishment and utilization of the mission objective
- Technical plan for implementing the mission with all mission elements
- Mission operations concept
- Management concept
- Rough time schedule for mission implementation
- Rough cost plan
- Financing concept.

A few papers and references do not describe the mission concept as part of the mission elements because of its nontechnical character.

8.1.1.3 The Launch Element

The **launch element** is the space transportation system for putting the space element into space. It consists of the rocket, including its upper stages, the launch pad, the launch facility, integration buildings, the launch control center, technical equipment, and facilities for the ground support of the payload and launcher during the launch and its campaign (see also Chapter 3). An important part of the launch element is the launch strategy: whether it is a question of a dedicated single launch, or a shared launch between two payloads, or a launch as auxiliary payload together with the main payload (so-called "piggyback" launch). This opportunity can be used only for small satellites with a total mass of less than approx. 150 kg. The main advantage is the relatively low launch costs (in the order of 15 000 to 20 000 euros/kg payload mass (22 000–29 000 dollars/kg)), because these costs are mostly paid by the main payload customer of the launcher. The main disadvantage is in the fixed orbit conditions determined by the main payload. For the selection of the launch element technical parameters like the environmental conditions for the payload on the launch pad and during the launch, geometrical requirements for the payload (footprint and shape of the payload, maximum dimensions), mass properties, mechanical, electrical, electromagnetic and data interface requirements between payload and upper stage of the launcher, and others must be defined in addition to the orbit insertion parameters and tolerances. Further important parameters of the launch element are access to or connection with the integrated payload, and the environmental conditions for the payload on the launch pad.

8.1.1.4 Orbit and Constellation

The **orbit** of a space mission is the selected ellipsoid or circular flight path around the Earth. There are several types of orbits, differing in some basic parameters, namely:

LEO – Low Earth Orbit
MEO – Medium Earth Orbit
GEO – Geosynchronous Earth Orbit
HEO – Highly inclined Elliptical Orbit (also called Molniya orbit).

Some characteristic parameters of the different orbit types are given in Table 8.1.1.

Classifying the orbits by functional criteria, they can be divided into:

- **Injection orbit**, the orbit achieved at the time of separation between upper stage and payload.
- **Parking orbit**, the orbit in which the spacecraft coasts temporarily without propulsion before it is reaccelerated to reach its destination orbit.
- **Transfer orbit**, the orbit or orbit segment used by the spacecraft to transfer from the initial orbit to a target orbit; for example, the **geostationary transfer orbit** (GTO) is a **Hohmann orbit** used to transfer a satellite from a low Earth orbit into a geosynchronous orbit.

	LEO	MEO	GEO	HEO
Average altitude [km]	300–1000	6000–25 000	35 786	500–36 000
Inclination [deg]	0–99	any	0	63.44
Orbital period [h]	≈ 1.5	≈ 5–12	24	12
Contact time [h]	≈ 10 min	2–4	24	11
Utilization	Miscellaneous	GNSS, communications	Communications, Earth observation	Communications

Table 8.1.1: *Types of Earth orbits and some characteristics.*

- **Target orbit**, the final orbit selected to fulfill the mission objectives by routine mission operations or destination orbit for orbital maneuvers.
- **Graveyard orbit**, the orbit for disposal of the spacecraft after the end of the mission. Every geostationary satellite has to leave its position after the end of the mission and move hundreds of kilometers upward, out of geostationary orbit.

A **satellite constellation** is a geometric arrangement of at least two satellites in given fixed orbits without any autonomous on-board orbit control and dedicated to the same mission objectives. Usually a constellation of identical satellites is implemented in space to increase the spatial and time coverage of the Earth. Satellite constellations are already established for communications (e.g., Iridium), for Earth observation (e.g., RapidEye, disaster monitoring constellations), for navigation (e.g., GPS, GLONASS and in the future Galileo) and for other objectives (e.g., military objectives).

A **satellite formation** is a geometric arrangement of at least two satellites in a given orbit or in relatively close orbits with a closed loop on-board orbit control. The satellite formation is dedicated to one mission objective. It is often called **formation flying** or a **satellite cluster** (e.g., GRACE).

8.1.1.5 The Space Element

The **space element** is the central element for achieving the mission objectives. It can be a single spacecraft or an arrangement of spacecraft in a constellation or formation. The space element can be a Space Shuttle, a reentry capsule, the space station or one of its modules, a satellite, or an interplanetary spacecraft. In Figure 8.1.1 the space element is a satellite. A **satellite** is defined here as an artificial unmanned celestial body orbiting the Earth. In contrast, an artificial unmanned celestial body leaving an orbit around the Earth is called a **space probe**. Both vehicles, having their own propulsion system, are also called spacecraft.

A satellite consists of the bus and payload. The **satellite bus** comprises all subsystems necessary to support the payload and the satellite bus itself in space. It keeps the payload mechanically stable in the desired direction relative to the bus structure; it puts the payload into the target orbit and keeps orbital position; it keeps the satellite in defined attitude modes and executes attitude maneuvers; and it supplies the satellite with electrical power, controls the thermal regulation, assures the reception, decoding and execution of commands, organizes the acquisition, recording and transmission of housekeeping data (telemetry data) from the subsystems and the payload, as well as the payload data. The **payload** of a satellite is dedicated to fulfilling the mission objectives. It could be a single scientific instrument, an instrument complex for intricate research objectives, a system of transponders for communication purposes, an instrument for operational weather observation, or any number of other options.

8.1.1.6 The Communication Architecture

The **communication architecture** describes the method of communication between the space segment and the ground segment and the arrangement of all the necessary elements. The communication architecture of a mission can include different communication modes on different communication paths, for example, voice communication in the UHF band, video communication in the S-band, telemetry and

telecommand data in the S-band and instrument data in the X-band. The communication architecture describes the interrelationship of all communication elements, paths and systems of a mission. The basic mission elements arranged in the communication architecture are the space element, the ground stations, the mission control center and the mission operations concept. In the architecture scheme the main data flows, such as the command, telemetry and payload data flows, are defined.

Different criteria can be used to define the communication architecture, for example according to the satellite–ground geometry, or the functions to be performed. Based on the data distribution principle, the satellite communication architecture can be distinguished as a point-to-point connection or as broadcast communication.

Point-to-point architecture is characterized by a link between the space element and one ground station, either directly or via relay satellite, for the transmission of commands, telemetry or science data.

Broadcast architecture is distinguished by the nondirectional transmission of data, telemetry or commands to many receiving stations, directly or via relay satellite. The transmitter might be one single ground station or a single satellite and the receivers might be many satellites or many ground stations.

Both of these architectures are a coarse classification of communication architectures that can be further classified in more detail according to the specific function. With reference to the main architecture classification, the following features can be distinguished:

- **Store-and-forward Communication:** Data or messages from one or several ground stations are received and buffered by the space element for a certain time (distance) and then transmitted to the target ground station(s) as a point-to-point connection or as a broadcast service.
- **Telemetry and Telecommand Communication:** The telemetry data of the space element is transmitted directly or via relay satellite to a single ground station (point-to-point) or to different ground stations (broadcast); the commands are transmitted from a ground station to a single satellite (point-to-point) or to several satellites (broadcast).
- **Data Collection Architecture:** Data is collected from the ground by satellite sensors or communication systems and transmitted directly or via relay satellite to a single ground station (point-to-point), or distributed via broadcast within the receiving range.
- **Data Relay Architecture:** A relay satellite transmits data from a ground station or from a satellite to one (point-to-point) or to several (broadcast) ground stations.

8.1.1.7 Satellite Ground Stations

Satellite ground stations comprise the transmitting and receiving systems on the ground for communication with the satellite. A satellite ground station consists of:

- A **controllable antenna** for receiving telemetry and payload data from the space segment and for transmitting commands.
- An **antenna control system** for orientating and pointing the directed antenna to track the satellite. The satellite tracking mode can be accomplished in two ways: "program tracking" uses calculated flight path data and "autotracking" uses the received signal power.
- An a**ntenna feed** with the radiofrequency part of the receiver or transmitter, including a low-noise amplifier (LNA) for received signals, and often including a downconverter within the receiving signal chain to convert the high-frequency signal into an intermediate frequency (IF) for transmitting the signal to the receiver unit with low signal losses.
- A **receiver unit** with a demodulator and sometimes with a bit synchronizer and a frame synchronizer for digital signals.
- A **transmitter unit** with high-frequency generator, modulator, upconverter and high-frequency amplifier.

If more than one ground station is used for one space mission, then that ground station will be defined as the **primary ground station** and the others as secondary ground stations.

The main ground station assures complete satellite control in the nominal case. The other ground stations receive the telemetry and payload data and, depending on the mission, assure the operation of

communication links. The ground stations are connected to the mission control center and to each other via a network. This could be the Internet, or it might consist of dedicated leased lines.

8.1.1.8 Mission Control Center and Mission Operations

The **mission control center**, sometimes called the space flight control center, is the focus for planning and operating the mission. It is the technical institution for monitoring and controlling the space element in space and time, as well as the appropriate ground elements and resources. It is responsible for assuring a stable link between the space element and the user or space service provider via a ground station or a ground station network. The mission control center of a satellite mission can be identical to the **space craft control center**, monitoring and commanding all subsystems of a spacecraft and controlling and operating the spacecraft bus in space and time. The third necessary element for the control of a satellite mission is the **payload control center**. It performs the same tasks as the spacecraft control center, but is dedicated to the payload of the satellite. It can be integrated within the mission control center and it can also be locally separated from the spacecraft and mission control centers.

An integrated mission control center, which includes the satellite and the payload control center, consists of the following main facilities:

- **Command System:** A database-driven system with all commands and command parameters and with a man–machine interface for mission control (control of the spacecraft and payload).
- **Telemetry System:** A database-driven system with all telemetry values and automated checks and limit controls and with a man–machine interface for mission control (control of the spacecraft and payload).
- **Archive System:** Permanent storage of all received telemetry data and transmitted commands.
- **Off-line Analysis System:** For the systematic evaluation and analysis of historical telemetry data according to preselected criteria, and for the analysis of trends related to the equipment, subsystems and payload.
- **Data Transmission System:** To maintain the link between all ground elements.
- **Mission Planning System:** To gather the operational user requirements and plan the mission operations according to these requirements, taking into account relevant constraints like visibility conditions for optical sensors, flight path parameters, energy budget of the spacecraft, available on-board payload data storage capacity, parameters of the next ground station contacts, and other factors.

The mission operations executed by the mission control center comprise the following tasks:

- **Mission planning** concerns all activities to collect the user requirements, to establish the time schedule for the operation of the ground station(s), to generate the command files, to support special requirements from users or mission control, to compute orbit and ephemeris data, to operate and use the links, and to analyze capacity utilization, among other things.
- **Mission operations** for the space element, which means commanding the spacecraft, monitoring the status of all subsystems, payload management, operation of the data acquisition systems of the spacecraft, reacting to spacecraft anomalies, and trend analysis related to the spacecraft's equipment and subsystems.
- **Instruction and training** of the operations team, including the development of training courses, training on simulators, training on spacecraft and further education.
- **Scientific and technological support** for the developer of the space element, especially during the design, implementation, integration and testing activities, planning orbital maneuvers, subsystem management, payload management, power management, reaction to anomalies, supervision of the flight software, supervision of the simulator (if available), maintenance of the databases of the ground station and mission control center, and trend analysis of the spacecraft.

8.1.1.9 Data Processing, Archiving and Distribution

Space missions to gather scientific, operational or administrative data need technical facilities for

systematic processing and archiving the data, so-called data **processing and archiving facilities** (PAFs), as well as technical facilities for data distribution. These facilities are centralized, and can be part of a satellite ground station. The users obtain the data from the data processing center after it has been processed to a standardized level.

8.1.1.10 The Users

The users of a space mission determine the mission design and its implementation with regard to given constraints, for instance funding. Users of space missions may have a variety of different interests. According to their social role and interests, users can be divided into the following categories:

- Scientific users
- Commercial users (e.g., communication companies)
- Operational users (e.g., weather services)
- Nonprivate industrial users (e.g., semigovernmental enterprises)
- Official users (government agencies, administrations)
- Military users
- Nongovernmental users
- Universities and other educational institutions.

8.1.2 Segments of a Space Mission

It is customary for those elements of a space mission which are in a close functional relationship to be combined into one segment. This makes it possible to divide the entire space mission into just a few major segments. With their help, the complex interrelationships of a space mission can be more easily and generally described. Often a space mission is divided into three segments: the launch segment, the space segment and the ground segment. But other arrangements are also possible. In [8.1.1] an additional program segment is introduced so that four segments of a satellite mission can be defined:

- **Launch segment**, comprising the launch system, the launch service and the necessary infrastructure.
- **Space segment**, comprising the development of all spacecraft equipment, the payload, the system engineering activities and tests.
- **Ground segment**, including the ground infrastructure, the equipment, the hardware and software for linking the space segment and the mission control center, and the data processing and archiving facilities.
- **Program segment**, including project management, project coordination, product assurance, interfaces, project organization and planning of resources.

This general division of a space mission into segments is modified depending on the character of the mission.

For example, the missions of the global navigation satellite system (GNSS) are divided into three segments. The **segments of a GNSS** are:

- **Space segment** with a constellation of navigation satellites.
- **Control segment** that provides and operates the system.
- **User segment**, consisting of the different user of the navigation signals.

The launch segment (the space transportation system) is completely disregarded here, because the segmentation was done under functional aspects from the users' point of view. The launch segment is needed only for implementation; it has no direct relevance to functions for the users.

Another kind of mission segmentation is done for Earth observation missions. Typical **segments of an Earth observation mission** are:

- The **space segment**, one or more satellites in orbit.
- The **ground segment,** comprising the complete infrastructure and the staff for mission operations. The ground segment consists of the main elements: one or more ground stations, the space mission control center, the mission operation facilities and the ground data system connecting the elements mentioned above. The organization of communications with the spacecraft is defined by the communication architecture and depends on the mission requirements and constraints.
- The **service segment** includes the user, as defined within the ESA GMES program, and consists of the facilities for ground data processing archiving and

distribution (the PAFs) and the users with their own technical facilities.
- The **launch segment**, which is sometimes mentioned separately, includes the launch strategy, the space transportation system, the launch pad, the launch campaign, and the necessary infrastructure like checkout labs, living quarters for staff, communication lines and facilities.

The program segment is usually not defined as a segment of the space mission, although it is an important element. It includes the mission objectives, program and constraints; the quality assurance philosophy; the technical and management plans; the time schedule; the expense and financing proposal; and basic reference documents (requirements documentation, utilization plan and others).

8.1.3 The Mission Architecture

The **mission architecture** describes how the main elements of a space mission are related. In Figure 8.1.2 the interrelationship between the main elements of the BIRD mission is depicted as an example. BIRD is a small satellite mission of the German Aerospace Center (DLR) for in-orbit verification of new technologies and for remote sensing of fires from space [8.1.2]. The mission is being accomplished exclusively with a German ground segment.

The architecture depicted in Figure 8.1.2 shows the different interrelationships of the space segment with the ground segment. The DLR ground station at Neustrelitz is a receiving station only, without any options to command the spacecraft. Systematic data processing and archiving of the BIRD payload data is also carried out in Neustrelitz (not depicted in Figure 8.1.2). The mission control center in Oberpfaffenhofen operates the satellite via the DLR ground station in Weilheim, which has both uplink and downlink facilities. The users consist of an international investigating team coordinated by the DLR in Berlin-Adlershof. This team arranges for test fires, for example, and then monitors them with remote sensing instruments on both aircraft and satellites for data validation and comparing to ground measurements made during a satellite overpass. The experimental ground station in Berlin-Adlershof demonstrates the possibility of early warning of fires with a tailored low-cost satellite receiving station that can be rapidly installed at any location on Earth.

8.1.4 Development of a Mission Concept and a Mission Architecture

The **mission concept** and the **mission architecture** show the interaction of the space mission elements. The development and planning of the interrelationships needed to fulfill the mission objectives are accomplished in successive steps. This process is described in the following using the small satellite mission BIRD as an example. The definition of the steps is as in [8.1.3].

8.1.4.1 The Mission Statement

At the beginning of a space mission the mission statement is drawn up. It is formulated in a few sentences only and contains three core statements:
- The motivation for the mission, which means the rationale and the initial situation.
- The mission statement or the mission idea, which means how the mission elements interact.
- The user or user groups of the mission.

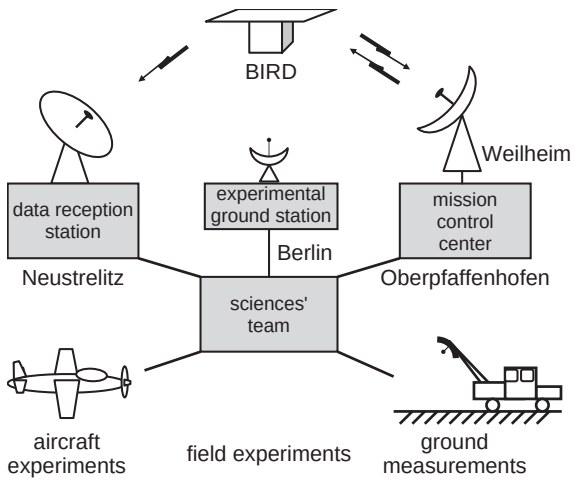


Figure 8.1.2: *Mission architecture of the BIRD mission.*

The mission statement should contain self-explanatory sentences and be clearly understandable to everybody.

Example of a Mission Statement
Because of the increasing impact of fires on forest and savannah ecosystems and on the atmosphere and climate, the global acquisition and measurement of fire parameters in space and time is of increasing importance. A dedicated satellite system with global coverage for daily acquisition and surveying of fire data in the regions concerned supports the daily briefing of those engaged in the management of major fires and the investigation of their implications by scientists, local authorities, organizations and insurance companies.

Such a spaceborne sensor system can also detect and support the remote sensing of other high-temperature phenomena like volcanic activity and can provide unique data for scientists and government administrators.

8.1.4.2 Definition of the Mission Objectives

The overall goal of the mission has already been formulated in the mission statement, which is a popular description. In a next step the mission objectives are precisely formulated, defined and described in detail. It is necessary that the customer, the mission developer and the foreseen mission operator develop a common understanding of the mission objectives. It can be useful to introduce a flat hierarchy into the mission objectives by a priority definition related to the different objectives (e.g., a division into primary and secondary mission objectives). The concept of the space mission is derived from the requirements and constraints of the mission. The requirements are derived from the mission objectives.

Example of the Definition of Mission Objectives (adapted from [8.1.2])
The primary mission objectives are:
- Test and space demonstration of a new generation of infrared sensors with an adaptive dynamic range for Earth observation
- Detection and scientific investigation from space of high-temperature events like vegetation fires, volcanic activities and coal seam fires
- Test and demonstration of new small satellite technologies.

The secondary mission objectives are:
- Test and demonstration of an on-board classification of remote sensing data using an artificial neural network classification method.
- Test and demonstration of an experimental ground station for immediate satellite data reception by the final user.

8.1.4.3 Definition of User Needs

Once the mission objectives have been formulated, the substantiated user needs are identified. Users can be roughly divided into three different main groups: scientific users, commercial users and governmental users. The needs of scientific users can be identified by calls for mission proposals or contributions. The needs of commercial users can be gathered by requests for proposals, though they are often more easily determined in user interviews with prepared questionnaires or at user conferences. The same methods are suitable for collecting and quantifying the needs of governmental users. A detailed analysis of user needs can lead to the modification of the mission objectives. After adjusting the objectives to the final user needs, the mission requirements can be derived.

8.1.4.4 Definition of Mission Requirements and Constraints

The mission requirements can be structured into functional and operational requirements. **Functional requirements** describe the tasks and essential performance parameters necessary to fulfill the mission objectives. The **operational mission requirements** describe the operational concept of the system for fulfilling the mission objectives and the interactions between the mission operations team and the user.

The functional requirements derived from the mission objectives can be grouped according to the segments.

Functional Requirements of the Space Segment
The functional requirements of the space segment are subdivided into the functional requirements of the payload and the payload platform, usually the

Table 8.2: *Examples of some functional requirements for a satellite remote sensing system (Source: [8.1.2]).*

Task	Spectral requirements	Radiometric requirements	Geometric requirements	
			Ground pixel size	Swath width
Fire detection and investigation	3.4–4.2 µm, 8.5–9.3 µm, 1 VIS/NIR	Infrared: DR > 2000 Resolution: 12 bits Saturation: T > 1300 K	> 300 m	> 100 km
Real-time cloud detection	min. 3 VIS/NIR channels + 1 TIR	DR > 1000 Resolution: 7 bits	< 1 km	> 100 km
Test of the on-board classification	3.4–4.2 µm, 8.5–9.3 µm, 1 or 2 VIS/NIR	Resolution: 7 bits (VIS/NIR), 12 bits (IR)	100–300 m	Low

VIS/NIR, Visible/Near Infrared; IR, Infrared; TIR, Thermal Infrared; DR, Dynamic Range; T, Temperature; K,Kelvin.

spacecraft bus. Payloads can be remote sensing instruments (for the scientific or operational observation of land, water, atmosphere, ice, Moon, planets, or other celestial bodies), scientific instruments for in-situ measurements (radiation, electric and/or magnetic fields, particle flows, and other phenomena), communication systems (transponders), satellite navigation payloads, or military payloads. An example of the formulation of the functional requirements of a payload for Earth observation is given in Table 8.2. Besides the requirements in the table, other requirements for optical payloads may also have to be taken into account, like stereo capability, different viewing angles, or the integer relationship between the ground pixel sizes of different instruments.

The first draft of the functional requirements document for the spacecraft bus results from the mission objectives and the user needs. The functional requirements for the spacecraft bus can be quantified with regard to the different functional tasks mentioned below. The spacecraft bus provides the following services for the payload:

- Transports the payload to the target orbit, performs the required orbital maneuvers, and keeps the payload in the required orbit or position.
- Keeps the payload mechanically stable and independent of internal and external disturbances.
- Points the payload with the necessary accuracy, keeps direction without jitter and avoids prohibited directions (e.g., pointing a telescope into the Sun).
- Supplies the payload with electrical power.
- Keeps the temperatures within the allowed ranges.
- Controls the payload so it can collect appropriate data.
- Acquires the housekeeping data of the payload and transmits it to the user.
- Other services, depending on the mission.

Further functional requirements concerning the payload and the spacecraft bus are the time and area coverage of the target regions, the payload and spacecraft modes, communication with the ground segment, the data processing algorithms and products of the payload data. The functional requirements for the spacecraft bus should be expressed quantitatively as early as possible.

Operational Requirements

The **operational requirements** for a mission concept concern the space segment and the ground segment. They result from the mission objectives, the necessary orbit, the concept for the space segment and the mission constraints. An important constraint for the definition of the operational requirements might be, for instance, the constraint to use only national ground stations.

Basic parameters for the definition of operational requirements are:

- Planned operational life span in orbit: for example, three years.

- Availability during the operational lifetime: for example, 95% for a one-year operational lifetime.
- Required orbit, for example:
  - Circular at 500–800 km altitude
  - Sun-synchronous inclination
  - Equatorial crossing time 10:30 local time.
- Required mission control center: for example, the German Space Operations Center (GSOC).
- Ground stations for the mission operations: for example, use of the national primary ground station only.
- Maximum duration of the program-controlled time-tagged operations of the space segment, for example:
  - Three days without ground contact
  - Duration of survival without ground contact; planned life span in orbit.
- Degree of autonomy of the space segment, for example:
  - Complete autonomy in the satellite's safe mode
  - Partial autonomy in nominal mode.
- Necessary ground operations activities, for example:
  - Arrangement and implementation of command sequences for three days minimum
  - Recovery from the safe mode, if it happens.

Functional Requirements of the Ground Segment
The functional requirements of the ground segment result from the mission objectives, the requirements for the satellite and the operational requirements.

The functional requirements of the ground segment are characterized by:

- Communication architecture: for example, store and forward.
- Ground station concept: for example, a prime ground station assures the bidirectional link (transmission of the telecommands, reception of the telemetry data) and several supporting ground stations receive telemetry and payload data and transmit it to the mission control center.
- Contact time with the ground stations.
- Data processing and archiving.
- Data dissemination.

Besides the functional and operational requirements for the mission, the constraints have to be taken into account.

Constraints
Constraints on the mission design are stringent demands for the accomplishment of the mission. In contrast to the requirements, they are not derived from the mission objectives but result from other economic, strategic, political and/or physical considerations. Constraints can be of very different types and should be distinguished from the requirements. Constraints can be for instance:

- The given development time for a space mission
- The given operational lifetime
- A fixed cost limit
- A funding model
- Cooperation with national, European or international partners
- The use of particular ground stations
- Compatibility with particular ground stations
- Cooperation with national entities or establishments and use of national technologies
- The space environment during the planned operational lifetime in orbit, especially with regard to the 11-year Sun cycle
- Others.

With the definition of the mission requirements and the constraints, the essential preconditions for the mission design are established. The next step is to design the mission.

8.1.4.5 Basic Mission Concepts and Alternate Architectures

A number of rough mission concepts should be developed to meet the requirements and constraints. These rough concepts will reveal differences in the mission elements and/or in the interrelationship between them. Different ways of relating the elements, for instance the space element and the ground stations, lead to different mission concepts. But all the defined mission concepts should fulfill the mission objectives under the given constraints. For a very strict set of constraints it might be possible

that only one mission architecture fulfills the mission objectives.

8.1.4.6 Identification of the System Drivers

System drivers are mission parameters which essentially determine the performance of the mission, its costs, the risks and the time schedule. The user or the mission developer can influence these drivers – they are not constraints. For each mission the system drivers should be identified and named.

8.1.4.7 Description of the Selected Mission Architecture

If various mission architectures have been identified, the criteria for the selection have to be explained and the preferable solution described. The interrelations of all system and mission elements needed to fulfill the mission objectives have to be pointed out unambiguously.

8.1.4.8 Identification of the Critical Requirements

In each space mission some requirements determine considerably more than others both the mission concept and the mission architecture. These critical requirements determine the system drivers. Some key questions for identification of the critical requirements are:

- Which requirements substantially determine the system design and which are the most difficult to fulfill?
- What kind of accuracy is required for each kind of task? What are the expected costs?
- What kind of operational requirements determine the system design and why?

8.1.4.9 Mission Analysis and Evaluation, Mission Value

The selected mission architecture has to be analyzed carefully with regard to compliance with the mission objectives. With the help of simulation tools the degree of compliance should be quantified and pointed out. If possible, performance criteria, figures of merit and measures of effectiveness should be used to judge the expected value of the mission.

8.1.4.10 Description of the Mission Architecture

The results of the mission analysis can lead to the modification of requirements and of the already selected mission architecture. Following the mission analysis the mission architecture is described and evaluated with regard to its expected benefit and value.

Bibliography

[8.1.1] IAA Position Papers, Subcommittee on Small Satellites: The Case for Small Satellites. *Acta Astronaut.*, **31**, 103–144, 1993.

[8.1.2] Brieß, K., Lorenz, E. Systematic Image Processing of the Small Satellite Mission BIRD. Proceedings of the 49th International Scientific Colloquium, Ilmenau, September 27–30, 2004.

[8.1.3] Larson, W.J., Wertz, J.R. (eds), *Space Mission Analysis and Design*, Third Edition. El Segundo, CA: Microcosm Press and Dordrecht: Kluwer Academic, 1999.

8.2 Systems Design and Integration

Klaus Brieß

8.2.1 Systems Design of a Space System

8.2.1.1 System Concept

A **space system** is a complex unit of elements which interact to achieve defined functions for a space activity. **Complexity** implies a high variety of elements and a variety of connections and relations between them. These elements can in turn be analyzed in terms of their constituents to achieve a higher level of detail. In the space sector the hardware terminology classification shown in Table 8.2.1 has been generally accepted.

Table 8.2.1: *Classification of space systems.*

System level	Annotations and examples
System	A complex arrangement of elements of a space mission. It may be divided into segments (space segment, ground segment, etc.)
Element	A complete functional system part consisting of subsystems. Examples are an instrument, a satellite bus, a launcher, a ground system, etc.
Subsystem	A unit of components and parts which form a functional subsystem, for example the structure and mechanisms, the thermal control system, propulsion, the attitude regulation system, etc.
Equipment	A complete functional unit, for example a battery, reaction wheels, a receiver, an antenna, etc.
Part	Circuits, screws, cables, materials, housings, etc.

The term **system** has several uses in the space sector. The connection of all subsystems, for example, forms the satellite bus. The action of connecting the payload module to the satellite bus is called "system integration" and it forms the system segment named "satellite." Several satellites which interact in space to complete a mission form a **satellite system**. This can also be described as the space segment of the mission. If there is interaction between different ground elements of a space flight mission the combination is termed a **ground system**. The integrated stages of a launcher form the **space transportation system**. So the term **system** is also used for segments as well.

8.2.1.2 The Design Process

The system design process for a space mission starts with an analysis of the requirements and constraints and a preliminary design of the system consisting of the mission elements. In the following step the elements are designed, and subsequently the subsystems and components. System design can be described as a "spiral" process with an increasing level of detail as it progresses. In the development process an inversion occurs in order to review and evaluate the designed solution against the requirements and parameters. This process is termed **system analysis**.

In the design process the following **principles of system design** should be noted:

- **Top-Down Design:** The design of a space system starts at the system level and proceeds onward to the elements, the elements' subsystems, the components, down to the level of parts, step by step, over the different project phases. The system has to be defined in its environment, and all interfaces must be specified at each system level.
- **Alternative Thinking:** At each system level different concepts and versions should be designed and evaluated in order to find the optimum solution to meet the requirements and constraints.
- **Scheduling in Phases:** Similar to structuring the contents (top down), the design process of a space system is chronologically structured in different phases, with the level of detail increasing step by step. The scheduled phase structure is adapted to the system life-cycle phases.
- **System Design According to a Fixed Procedure:** A system design process is executed according to a determined basic scheme. The main steps are: mission idea, situation analysis (demand, requirements, constraints), objectives, design of mission elements, analysis and evaluation of the different versions, decisions, specification of the mission concepts. The basic scheme of a design at system level is described in Section 8.1.

In general there are sequential tasks in a system design process, which are completed with a formal review procedure. In accordance with [8.2.1] the following tasks are defined:

1. **The Mission/Functional Activities:** These involve specification of the required mission and the functions. The formal end of the task is formed by the **preliminary requirements review** (PRR), which identifies the objectives of the mission and its feasibility. The associated system configuration is the so-called **functional baseline configuration**. The system solution can still be designed in different versions.
2. **The Requirements Activities:** Starting from the mission objectives, the requirements for the mission, the systems and elements as well as the

interfaces are defined and evaluated. The requirements are approved in a formal **system requirements review** (SRR). According to the approved requirements the system can be defined and specified on a top level. This task is concluded with the **preliminary design review** (PDR). The system configuration on this level is the so-called **development configuration baseline** of the system. This stage is called the "specified state."

3. **The Definition and Justification Activities:** These begin after the system requirements review. On the basis of a clear set of requirements, the system specification, the definition and explicit identification of the system are initiated by means of data, drawings and other technical documentation. They are the preconditions for the development, assembly, integration, operation and, if necessary, the maintenance of the system or element. The documentation for the explicit definition of the system (e.g., a satellite) is analyzed and examined in the **preliminary design review** (PDR). Afterwards, the technical documentation is refined and elaborated down to the level of equipment in an increasing level of detail. The system design is justified by detailed analyses, comparisons, evaluations, documentation or design reports and is terminated with the **critical design review** (CDR). The system is now ready for production (**production configuration baseline**) and this state is called the "defined state."
4. **The Verification Activities:** After the definition of the development configuration baseline (see task 2 above), the terms of assembly, test and qualification as well as the resources and methods of manufacturing, integration and verification of the system and elements are defined. The verification activities include analysis, test of development models, and qualification. The qualification process includes the theoretical and experimental verification that the defined system or element complies with the requirements and that it can be produced. The qualification also includes the resources and the process of production. The formal kick-off is the start of test activities at the system level: the **test readiness review** (TRR). The formal end of these activities is marked by the **qualification review** (QR), which leads to the "qualified state" of the system or element.
5. **The Production Activities:** The assembly and manufacturing activities start after the critical design review. Production covers all activities of assembly, integration and verification of the system. The product state and the test results of each model are documented in the **system acceptance review** (AR). The customer confirms that the model of the system complies with the defined baseline configuration. This stage is called the "accepted state." The acceptance review can also be carried out as a **preshipment review** (PSR) or **flight readiness review** (FRR).
6. **The Utilization Activities:** These relate to various user activities and correspond to the routine operational phase of the space system. These activities can start immediately after the launch of the satellite or with a delay until after the so-called commissioning phase. The formal beginning of this activity can be defined as being identical to the **operational readiness review** (ORR) or flight readiness review, or with the **launch readiness review** (LRR), which evaluates all activities related to the launch. The formal start of utilization activities can be defined as the beginning of operational activities. After liftoff and separation of the space element, the commissioning in space, the flight qualification and acceptance tests, and the measurement and evaluation of the performance parameters and the achieved flight level take place. This phase of putting the system into operation, called the **commissioning phase**, is formally closed with the **flight qualification review** (FQR). Thereafter, the system can go into routine operation and routine utilization activities of the system can start.
7. **The Disposal Activities:** These activities include all procedures from the end-of-life status to final disposal, for example by a targeted reentry into Earth's atmosphere or by disposal in a graveyard orbit. This phase is prepared during the utilization phase and is specific for each system.

The coherence between these mentioned reviews in a space project and typical project phases is shown in Figure 8.2.1. This is a typical time schedule scheme for space flight projects.

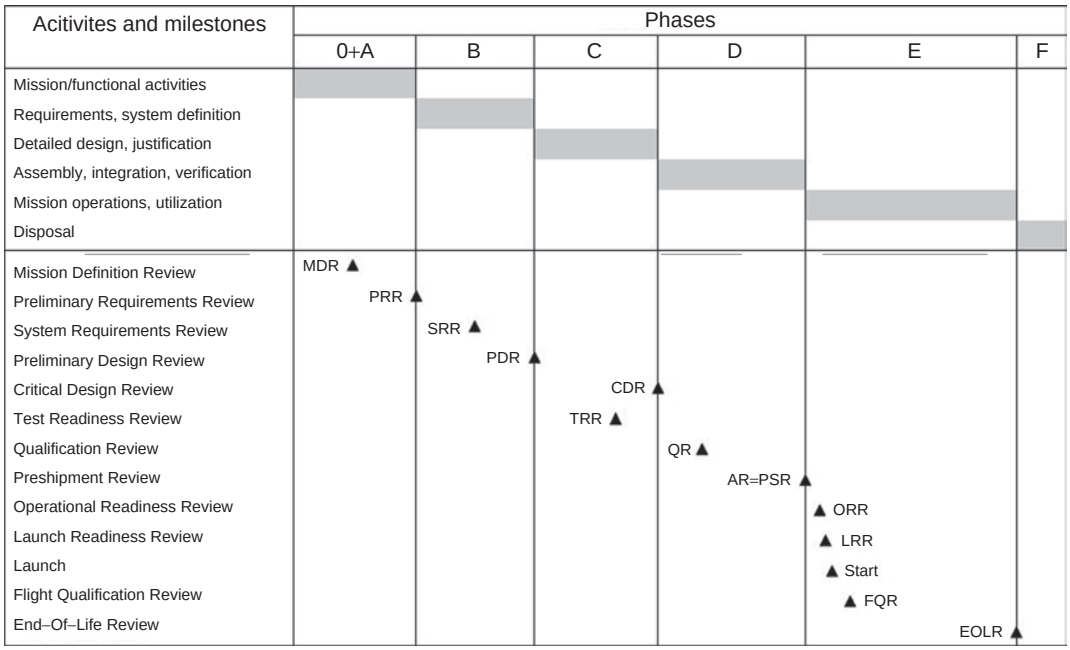


Figure 8.2.1: *Project phases and formal review procedures of a typical space flight project (following [8.2.1]).*

8.2.1.3 The Design Philosophy

The system design of a space mission is based on a design philosophy. Generally two types of design philosophies can be distinguished:

- Design to objectives
- Design to budget.

The design of a space system is carried out according to one of these two philosophies. This means that either the system is primary designed to achieve the scientific, commercial, governmental or other objectives, or the primary requirement of the system design is compliance with a given fixed budget.

Design to Objectives

The design philosophy of **design to objectives** can also be called **design to science** in the case of science missions. A design-to-science project focuses on completion of the mission objectives without any exceptions and cutbacks. The conception of all elements and segments of the space flight mission is justified by the achievement of the scientific, operational, commercial or military tasks. Completion of the defined mission objectives and tasks takes priority over the observance of a cost limit. This does not mean that costs are not taken into account, but that cost planning results from the objectives and the planned implementation of a well-founded mission concept. The cost plan has to be supplemented with an adequate financing concept. If the planned time and cost budgets exceed the limits during mission development and accomplishment, opportunities for an increase in the budget as well as personnel and technical resources can be examined and exploited. Cutbacks or a decrease in the level of completion of the mission objectives is practically unacceptable.

The design process according to the design-to-science philosophy is shown in Figure 8.2.2 using the example of a small satellite mission. The mission objectives and constraints evolve from the idea for accomplishing the mission. The budget is a typical constraint which must be considered when the different mission options are evaluated, but the ultimate design process is defined by the objectives and the functional and operational requirements of the mission. In Figure 8.2.2 the determining design flow is marked by the heavy arrows. The functional and operational requirements alone determine the

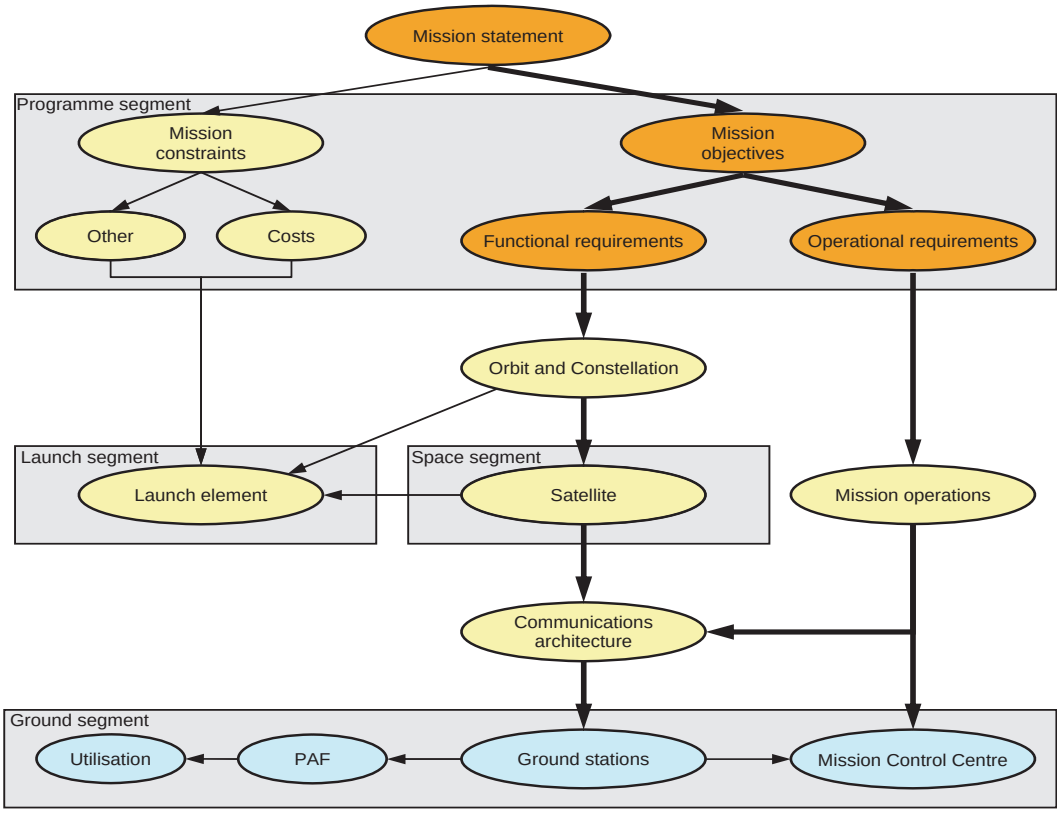


Figure 8.2.2: *System design scheme according to the "design-to-objective" philosophy. The main design flow is marked by the heavy arrows. (PAF: Processing and Archiving Facilities).*

elements of a satellite mission. In the given example the choice of orbit and the need for a satellite constellation result only from the functional requirements. The costs which result from the orbit selection are of secondary importance. The space segment, which means the satellites and their payloads, is defined by the functional requirements and the optimum orbit or trajectory. The operational requirements define the mission operations concept. The costs which result from the mission operations come second in this case too. The communication architecture, the ground station concept and mission control result from the mission operations concept. This process defines how many times the satellite has to be contacted and how many times it has to dispose of its data or has to be loaded with new command sequences, and at what intervals it has to be monitored and controlled.

Design to Budget

The **design-to-budget** or **design-to-cost** philosophy makes compliance with the budget limit the arbitrative criteria for the design, development and mission performance. Completion of the mission objectives is in focus during the design process, but not without cutbacks and not at any cost. The design of the elements and segments of the space mission is focused on fulfilling the scientific, operational, commercial or military tasks and objectives, but only to the extent that the budget is not exceeded. If a mission is not feasible within the existing budget, the objectives have to be reviewed by identifying the design drivers, then iterated and modified or reduced, if necessary. If neither option is feasible, the mission cannot be accomplished.

The approved budget determines the design of all elements and segments of the mission right from the beginning of the project. Figure 8.2.3 shows a

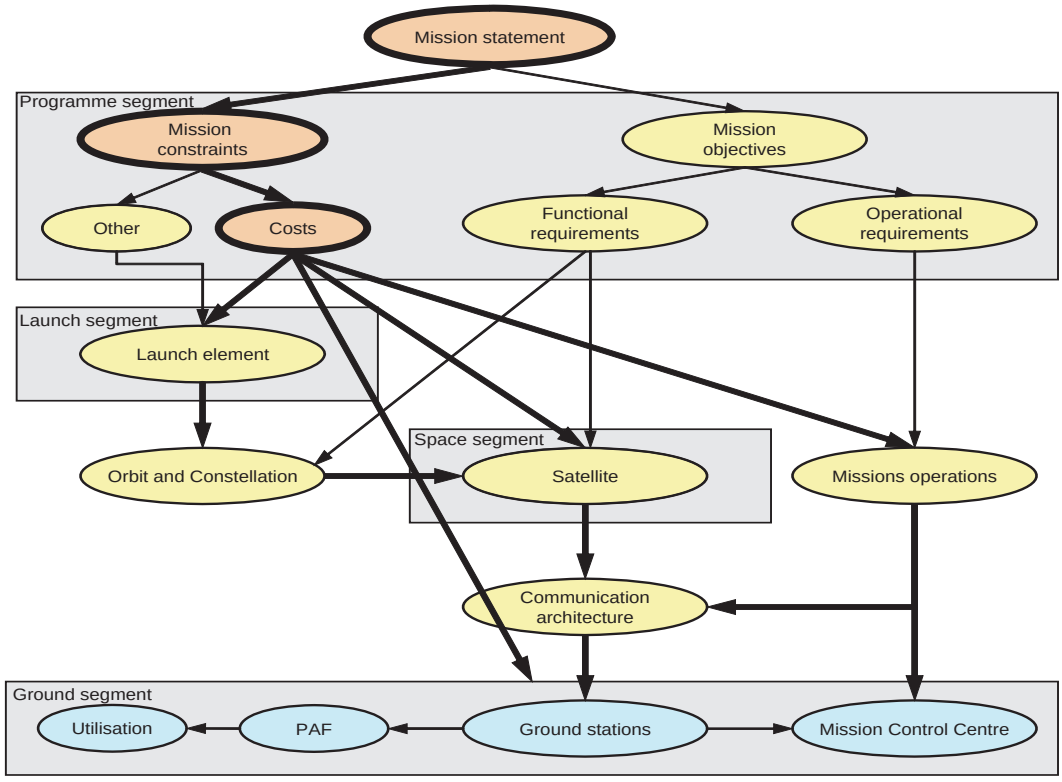


Figure 8.2.3: *System design scheme according to the "design-to-cost" philosophy. The main design flow is marked by the heavy arrows (PAF, Processing and Archiving Facilities).*

schematic design process according to the design-to-budget philosophy. The design of the mission elements is exclusively defined by the cost budget. For that reason the operations concept cannot be primarily designed according to the operational requirements. This can lead to providing only limited support in the ground segment, for example. Costs play a decisive role in the definition of the launcher, since it is usually one of the cost drivers of the mission. For the first time questions arise about the launch strategy: Can there be a dedicated launch to the desired orbit, or is there an opportunity for a satisfactory compromise with other missions by arranging for a "shared launch" in order to split the launch costs with a partner, or an opportunity for a "piggyback launch" (meaning that the satellite can be carried as an auxiliary payload)? The answer to these questions defines in the first instance the orbit and the space segment. After this decision, the design of the other elements can be derived.

8.2.2 System Integration

8.2.2.1 The System Integration Process

System integration is the process of assembling and integrating the subsystems and equipment. This results in the completion of the space system, which is one of the system segments. Usually, the process of integrating the satellite bus with the payload is called system integration. Highly integrated systems are, however, not connected with an already completed satellite bus; instead, the satellite bus is constructed around the payloads. Therefore, system integration also means the assembly and integration of all subsystems of the bus and payloads. The system integration process is individually tailored to each satellite. For this reason there is no generally valid definition of the process.

Typical features of the system integration process are:

- All subsystem and payload development has been completed. All components, materials, tools and devices which are necessary for system integration are available, or their delivery is definitely assured in time for integration.
- System integration requires auxiliary facilities for handling, assembly and tests. This equipment has been developed separately and is available in time.
- System integration takes place in an integration room with restricted access.
- The steps of system integration and the necessary tests for system verification are specified in an operations chart.
- The integration process is divided into stages which are completed with a review and test of system status. This process is called **system verification**. The power subsystem, for example, is verified after the integration process. All components and their connections with the computers are tested after integration of the on-board computers.
- All steps are documented in an accompanying document.

In the definition of the integration sequence two aspects have to be considered:

- The function
- The access.

This leads to a commonly accepted sequence: the mechanical baseline structure is built by using the mechanical ground support equipment. Depending on the accessibility, components of the thermal control system are also integrated (heat conductors, radiators, etc.). In the case of a satellite with a propulsion system, the main components of the propulsion system are integrated (engine, nozzle, tanks, valves, pipes and other equipment). Then the wiring and, depending on accessibility, further components are integrated. Because of its function, the first electrical system to be integrated should be the power system. In the next step the on-board computer is integrated, followed by the attitude control system and the communication system. Depending on access, the payloads are integrated. In the case of integration of an optical payload, the payload can be verified after integration by optical ground support equipment (OGSE). As mentioned above, the integration process is an individualized sequence and can vary from satellite to satellite. In any case, ground support equipment is necessary for system integration and test.

8.2.2.2 Integration and Test Facilities

System integration takes place in an integration room with well-defined environmental conditions. Such conditions concern, for example, the temperature range, relative humidity, vibrations, EMC, etc. A measure for the dust or particle density in the air is the clean-room classification. The **clean-room class** specifies the number of dust particles with a diameter ≥ 0.5 µm per cubic foot of air. Whether a clean room is necessary or not has to be decided from case to case. A clean room is a laboratory room where contamination is controlled. It is not necessary for small satellite projects not requiring any special work on optical surfaces, open electrical integrated circuits or optical sensors. In this case, adherence to typical laboratory conditions is sufficient:

- Temperature: 20 ± 5 °C
- Relative humidity: 50%
- Dust particles in typical downtown concentrations, particles < 5 µm (smoke, ash, dust) are floating in the air.

A typical value for the clean-room class where technical equipment provides air cleanliness is 100 000. For satellite integration clean rooms, values of 10 000 can be required. For satellites designed and developed by universities a clean-room class of 200 000 is sufficient. This is in accordance with clean standard laboratory rooms without technical air cleaning equipment.

For test support during and after system integration, access to the following support equipment and facilities is required:

- Measuring facility for analysis of mass properties
- Vibration test facility
- Acoustic test chamber (if necessary)
- Thermal chamber
- Vacuum chamber (or thermal vacuum chamber)
- EMC test facilities (if necessary)
- Magnetic field measuring facility (if necessary)
- Attitude control test facility (if necessary)
- Ground station for compatibility tests.

8.2.2.3 Ground Support Equipment

For the integration process, ground support equipment is necessary to transport the space system, to enable access from all sides, for the wiring, for electrical stimulation and testing, and for optical stimulation

and testing, if necessary. The required ground support facilities at system level are:

- **Mechanical Ground Support Equipment (MGSE):** The MGSE lifts, holds, turns and transports the system during the integration process. It enables access to the system from any direction. The MGSE has to be verified for these applications. It consists of different parts, for example a system support unit, a crane, a forklift, a fueling unit, etc.
- **Transport Container:** The transport container is special mechanical support equipment for the safe transport of the space system from the integration facility to the launch site or test facilities, as required. It is a hermetically closed container with attachments for the space system, vibration dampers, shock absorbers, shock recorders and humidity absorbers. After closing it should be able to be flushed with nitrogen.
- **Electrical Ground Support Equipment (EGSE):** The EGSE is a test facility for electromechanical and electrical components, equipment and subsystems of the space system. It applies a voltage to the components and parts and enables the generation of test signals, receives data from the tested system, records the data and displays the test results immediately. It is used for complete electrical verification and allows the complete checkout of the space system, whether a transponder, an instrument or a satellite. The EGSE is connected with wires and electrical interfaces to the space system. A satellite EGSE should be capable of communicating wirelessly with the satellite via the defined radiofrequency link.
- **Optical Ground Support Equipment (OGSE):** The OGSE stimulates optical instruments, records the response data and shows the results in a laboratory or at the launch site. The laboratory OGSE serves as a measurement facility for recording the geometric, radiometric and spectral calibration data of the optical instruments. From case to case it can also be designed to measure additional parameters, for example polarization sensitivity, stray light suppression, etc. The use of robust and mobile OGSE can also be required at the launching site to measure and test possible changes in the instrument characteristics caused by transport.

8.2.3 System Verification

As mentioned in Section 8.2.2, the system integration process is associated very closely with system verification. Integration is completed with the completion of verification.

8.2.3.1 The Objectives of Verification

Verification is the proof that the space system meets the requirements and is in accord with the required project life cycle [8.2.2]. The contractor is responsible for verification.

According to [8.2.2], the objectives of verification consist of qualification of the design, confirmation and proof that the product complies with the qualified design and is without manufacturing failures, confirmation that the space system and the mission operation staff (including tools, procedures, resources, etc.) are able to comply with the mission objectives and requirements, and proof of the performance parameters of the space system.

8.2.3.2 Stages of the Verification Process

The verification process is accomplished in different sequential phases according to the life cycle of a space mission project. The different verification phases according to ECSS-E-10-02A (1998) [8.2.2] are:

- **Qualification:** The objective consists of demonstrating that the designed system meets all the requirements, including margins. The qualification object must comply completely with the "full flight design" and the "flight standard" (e.g., QM, FM and PFM). The proof results from testing the mentioned models with test stimulations that are much more intensive than the acceptance stimulations.
- **Acceptance:** The objective of this stage consists of demonstrating that the product has no manufacturing and integration defects and that operational use is appropriate. The proof results from testing with stimulations which are a little more intensive than the expected excitations.
- **Prelaunch Verification:** The objective of the prelaunch verification stage consists of proving that the space system is ready for launch and early space

operations. Verification results from the outcome of tests and analysis.

- **In-Orbit Verification:** The objective of the "in-orbit" verification phase consists of confirming that the space system is qualified for the defined application in space. The evidence results from in-orbit tests, supplemented by ground tests.
- **Postlanding Verification:** The objective of the postlanding verification phase consists of verifying certain functions and system status after the mission. This is only valid for reentry systems.

8.2.3.3 Verification Methods

Verification is achieved by one or more of the following four methods [8.2.2].

Test: This is a method to verify the requirements by measuring product properties or functions under certain simulated conditions. A test can also include the demonstration of operational properties and requirements. The following kinds of tests can be distinguished:

- **Development test** for the analysis of new developments and demonstration of the suitability of new design concepts.
- **Qualification test** for the demonstration and evidence of flight acceptability of the designed system under the required launch and space conditions. A model in flight standard is tested with higher than expected stimulations and impact durations.
- **Acceptance test** to prove that the flight model complies with all requirements and is free of manufacturing defects. The stimulations are similar to the expected loads, but the impact durations are shorter.
- **Analysis validation test** for the acquisition of data under a strictly defined test environment to validate or improve mathematical models by testing them in the flight standard with low stimulations and short impact durations.

Analysis: This is a method of theoretical or empirical verification by evaluating properties using commonly accepted techniques. These analysis techniques are systematic, statistical and qualitative methods of analysis as well as simulation tests or verification by similarity of space systems.

Design Review: This method is characterized by using validating data, design documents, technical specifications and other plans for verification. This documentation must show that the design complies with the requirements.

Verification by Inspection: This is a verification method that confirms the conformance of the hardware and software with the relevant documentation (test reports, protocols). It provides visible evidence of the physical properties and condition of the object being verified, without the use of special laboratories.

8.2.3.4 Verification levels

Verification takes place at all levels with space mission hardware in accordance with Table 8.2.1.

8.2.3.5 Models for Verification

Verification of a space system is performed by using different models. The number of models is optimized with regard to the costs, risk and complexity of the verification.

Mock-up (MU): An MU is a model for the optimization and control of interfaces, for validating the integration process, for the accommodation analysis, the architecture analysis, human factor evaluation and the evaluation of operational procedures. It represents the geometric configuration, layout and interfaces. It is used above all for human factor evaluation and astronaut training, and similar tests.

Development Model (DM): A DM is a model supporting the development process and used to validate the feasibility of the design. It represents only the selected functions to be tested, for example mechanical, electrical or thermal functions. Size, shape and interfaces do not need to be representative, if they are not the test object. The DM is used for development tests at any verification level, especially at the equipment level.

Table 8.2.1: *System levels and verification methods.*

System level	Verification method
Part	Review, inspection
Equipment	Review, inspection, test
Subsystem	Analysis, test, inspection
Element	Analysis, test, inspection
System	Analysis, test, inspection

Integration Model (IM): An IM is a model for functional and interface tests; it is used to analyze failure modes and software, and to validate procedures. It represents the full functionality of the hardware and software. It is realized with commercial parts. Missing components are simulated. It is also called the "electrical model" and can be used at any verification level.

"Suitcase" Model: This is a model for the complete simulation of the communication system in compatibility tests with the designated ground stations. This includes interface tests and failure mode analysis. It represents the complete functionality of all radiofrequency equipment in the communication system (transmitter, receiver) of the flight model across all parameters and it contains the necessary functional simulations like decoder, transponder, etc. The use of commercial off-the-shelf components is allowed. It is used for qualification and compatibility tests with the ground segment.

Structural Model (SM): This is a model for the qualification of the structure and validation of the finite element model of the structure. It represents the flight standard regarding the structural parameters. It is used for qualification tests.

Thermal Model (TM): This model is dedicated to the qualification of the thermal design and to the validation of the thermal mathematical model. It represents the flight standard regarding thermal parameters and properties. It consists of thermal dummies. It is used at the subsystem and system levels (thermal control system).

Structure and Thermal Model (STM): This is a combined model of the TM and SM (dummies) for cost-saving reasons. It is used for qualification tests at system level.

Engineering Model (EM): The EM is used to qualify the functions, to verify the fail-safe system operations, and for testing important parameters. It represents the electrical functions by using commercial components and is built in a flight-typical configuration. It represents the flight design without redundancy and high-reliability parts. It is used at any system level.

Qualification Model (QM): This model is used for qualification of the design of a component or subsystem of a space system. It represents the complete flight design and standard. It is used at the component and subsystem levels or for qualification tests.

Engineering Qualification Model (EQM): This is a model for the functional qualification of the design and the interfaces, as well as for the verification of electromagnetic compatibility (EMC). It represents the complete flight design without the high-reliability components, which can be substituted by military standard (MIL) parts of the same manufacturer. It is used at any system level for functional qualification tests.

Flight Model (FM): This is a model for space application. It represents the complete flight design and flight standard. A flight model is built at any system level. It is tested at the acceptance level only.

Protoflight Model (PFM): This is a model for space application and for qualification of the design. It represents the complete flight design and flight standard. It is used at any system level. The PFM undergoes qualification tests and can be used for space application after completion of the qualification procedure.

Flight Spare (FS): This model is consistent with the requirements for space applications. It represents the complete flight design and standard. It is used at the component level or the system level for payloads and small satellites.

Further Models: Apart from the models mentioned above, additional models can be developed for the verification process, for example a functional model to qualify or validate functional and operational concepts, to verify mass properties, or to train staff.

8.2.3.6 Model Philosophy

The number and kinds of models used for system verification depend on the selected philosophy. The philosophy is defined by the mission requirements and constraints. Generally the following three model philosophies can be distinguished.

Prototype Philosophy: This is a model philosophy characterized by extensive use of different models for system verification (note that it requires many prototypes). This approach reduces the technical risk. The philosophy is used in the design and development process of very new and complex systems, interplanetary space flight missions and for special requirements. The main disadvantage is in the high total costs which result from the manufacture and use of many models.

Protoflight Philosophy: This is characterized by the use of only one model, which flies after the qualification and acceptance tests. The use of only

one model reduces the costs to a minimum. In order to limit the risk, the philosophy is only applicable to systems without critical technologies and to systems mostly using qualified hardware.

Hybrid Philosophy: This philosophy is a compromise between the prototype and protoflight philosophies, tapping the advantages of both philosophies. The objective is to develop new and complex systems with as few models as possible by keeping the total risk to an acceptable limit. A protoflight model is implemented in space after a reduced protoflight qualification. New and critical technologies are tested in separate qualification models at the qualification level. For that reason only acceptance tests are executed on the protoflight model at system level. For example, if the structure according to the hybrid philosophy is already qualified by an STM, the structure of the protoflight model will only be tested at the acceptance level.

8.2.3.7 The Hardware Matrix

The hardware matrix shows systematically line by line all the necessary equipment of a space system arranged according to the subsystems. After the model philosophy has been defined, all necessary models of the equipment are charted in the columns. A classification with regard to the qualification status of the equipment makes it easier to define the required models. The qualification status is divided into four classes (A, B, C and D). "A" means already qualified off-the-shelf equipment; no further qualification is necessary. "B" means small variations requiring a delta qualification. "C" means more extended variations of the equipment requiring either a delta or a complete qualification. "D" means a completely newly developed or extensively modified component which requires a complete qualification. Table 8.2.2 shows an extract of a hardware matrix for a satellite as an example.

8.2.3.8 The Verification Matrix

The system verification strategy in the different requirement categories is defined in the verification matrix. The matrix lists systematically line by line the different requirement categories and in the columns the different system levels (system, subsystem, equipment). The verification methods are charted in the matrix arrays (Table 8.2.3).

Table 8.2.2: *Extract of a hardware matrix for a satellite (example).*

Subsystem/ instrument and equipment	Qualification status	DM	STM	EM	FM
Structure	D		1		1
Thermal control	D		1		1
Power					
Power control unit	A		1*	1	1
Battery charge unit	B		1*	1	1
Battery control unit	C	1	1*	1**	1
Power distribution unit	D	1	1*	1**	1***
Battery	A		2*	2	2

*Dummy, **EQM, ***PFM.

Table 8.2.3: *Extract of a verification matrix of a satellite (example).*

Requirement category	System	Subsystem	Equipment
Structure			
Mechanical–dynamic	T, A	T, A	T, A
Mechanisms and pyrotechniques	T, A	T, A	A, R, I, T
Thermal control			
Thermal control system	A, T	A, T	A, T
Power			
Power generation	A, R, I, T	A, T	A, I, T
Power control	A, R, I, T	A, T	A, R, I, T

A, Analysis; T, Test; I, Inspection; R, Review of design.

Bibliography

[8.2.1] ECSS-M-30A. *Space Project Management – Project Phasing and Planning.* Noordwijk, 1996.

[8.2.2] ECSS-E-10-02A. *Space Engineering – Verification.* Noordwijk, 1998.

8.3 Environmental Tests and Basic Concepts

Ralf Baumgartl, Christian Henjes, Ulf Hülsenbusch, Anton Grillenbeck, Holger Kügler, Wilfried Ley and Alf Schneider

8.3.1 Principles and Significance of Environmental Tests

The two concepts of environmental and functional testing are the basis of almost all **development and acceptance tests** (simply called "tests") of space equipment. Environmental testing means to provide, to approximately simulate or to directly apply the environmental conditions that occur during launch or during the mission to the space equipment under test.

The **environmental conditions** in this context comprise all natural or artificial environments that can possibly have an effect on the space equipment's life cycle. The most important but not the only environmental effects or environments are:

- **Mechanical:** Static, dynamic and aerodynamic loads, vibrations and shock
- **Thermal:** Vacuum, internal and external thermal sources and sinks
- **Electromagnetic:** Artificially generated, natural and sometimes high-energy processes and magnetic fields
- **Special Environmental Effects and Conditions:** Plasmas, microgravity, impact of micrometeorites, near-Earth or interplanetary space, specific orbit conditions (for instance, the radiation belt surrounding Jupiter).

Due to the complexity and diversity of environmental conditions, environmental tests are preferably conducted separately in test programs, even if the conditions exist in parallel. A combination of environmental conditions in a test is not considered an option unless operational restraints necessitate this approach, such as thermal and mechanical conditions on entry into an atmosphere or the operation of an engine during atmospheric ascent.

Functional tests, in contrast, are dedicated to proving the reliable behavior of space equipment during its operational life. It is evident that a wide variety of functional tests exist.

In this section of the book, the focus is on environmental simulation, sometimes in combination with functional tests, as required. One of the main reasons for this focus is that **environmental simulation tests** for space engineering are in general understood and implemented as a standard for development and acceptance tests. Moreover, environmental testing requires considerable effort to provide adequate test facilities and capabilities. Accordingly, environmental tests for space engineering are performed exclusively in very specialized test centers. In this context this section deals with the experimental investigation of the physical and technical properties of space equipment, since these properties are of particular importance for equipment operating under the environmental conditions of space. Examples are investigations of dynamic properties (rigid body properties and modal parameters).

The objective of environmental tests within the scope of development and acceptance testing is to demonstrate beyond doubt and on the ground the reliability of space equipment under the anticipated (space) environmental conditions. Furthermore, it has to be proved that the operational behavior of the space equipment can be kept within the specified tolerances during the planned mission period in the specific environments. In the following sections, it will be shown that checks and environmental tests are carried out not just at the end of the assembly phase. During the complete development phase of space equipment, environmental tests are systematically planned and performed in order to achieve complete proof of design **suitability for space applications**.

The proof of suitability is usually demonstrated in two stages: **qualification** and **acceptance**. These two stages differ as to the number of tests involved as well as their extent and intensity:

- **Qualification Test:** Formal experimental verification that the realization of the design and the manufacturing procedures result in space equipment which meets the requirements or specifications. The applicable test requirements surpass the respective flight requirements including margins,

even in cases where all tolerances coincide in the most unlikely configuration.

- **Acceptance Test:** Formal experimental verification that the space equipment is free of manufacturing defects and integration errors and is ready for its intended use. The anticipated maximum flight requirements can be fulfilled without restriction or diminished functional capability.

Several standards have emerged for the planning and performance of environmental tests for space flight. The most important within ESA are ECSS-E-10-03 (Space Engineering – Testing) and ECSS-E-10-02 (Space Engineering – Verification), [8.3.1] and [8.3.2] respectively. For NASA, the series of standards NASA-STD-5000 and 7000 are applicable [8.3.4]. These standards are understood to be the baseline for any verification and test program. However, adaptations are possible in principle; that is, they can be tailored to new or altered mission requirements, technological innovations or developments. Accordingly, in any specific case the test effort may be increased or reduced.

In order to show the significance of environmental tests, in particular with respect to tailoring, a few basic insights and developments are listed here:

- The success of a space program is not simply the result of the strict application of standards. Much more decisive is the testing of the built space equipment under appropriate conditions, and a guarantee that the verified equipment will be used as verified. Simply said: "Test what you fly, fly what you test." Some failed space programs which were designed in the era of "faster, better, cheaper" have forcibly brought this lesson home.
- There are, however, a vast number of successful or less successful space programs. Within the framework of the ESA-funded initiative MAT D [8.3.5], an impartial check for failures in many of these space programs led to interesting conclusions as to the effectiveness of specific tests. Reasons to perform or not to perform a specific test in order to reduce risks can be derived using statistical means. As more detailed results and conclusions of this initiative are beyond the scope of this handbook, reference is made to the literature.
- Besides the purely technical aspects of checks or, in particular, of environmental simulation tests, the shared production of space equipment requires qualification and acceptance tests as a basis for the acceptance of subcontracted work.

In Section 8.2 the purpose of a verification matrix is discussed. It forms the basis for a **test matrix** describing environmental and functional tests in detail by indicating the type of test and specific test requirements. Finally, the elaboration of test requirements for a specific test or even a series of tests is laid down in a **test plan**. In this plan, the test objectives, the test methods, the test article configurations, the relevant (environmental) conditions or test loads, the test article functions to be investigated, the safety requirements to be observed when handling the test article, and the physical or electrical parameters of the environment or test article to be measured or monitored are described in detail. Moreover, the criteria defining success or failure of a test are listed. A **test procedure** is prepared following the guidelines for the test plan for each individual test; it includes detailed step-by-step instructions on what parameters are to be measured with the test facility, how the test is to be performed and when measurements are to be taken.

The sequence of tests within a test plan usually follows two principles. The first is that the sequence of environmental conditions and functional operations should reflect actual operational life. The second is that potential failures and problems should be detected as soon as possible within a test program to avoid the cost of repeated testing.

In order to ensure the last aspect, a **model philosophy** is designed for each test program which assigns physical test articles to the various tests. The objective of a model philosophy is, as already discussed in Section 8.2, to define the optimum assignment of test article properties and the number of test articles for the individual tests. This should be done so that highly efficient and reliable product verification can be achieved in a short time and with reasonable consideration of costs and risks.

8.3.2 Verification Planning and Cost Factors

Planning the **verification and testing** procedures for a space project is just as complex a process as the design

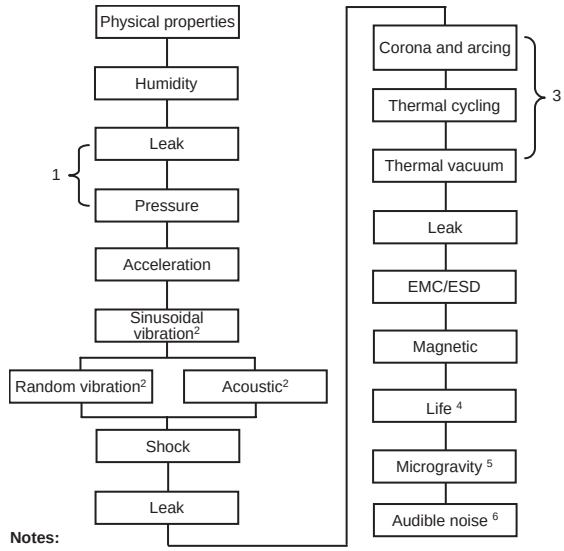
and implementation of the space equipment itself. Moreover, this process is highly dependent on decisions made during the preceding phases. This must by all means be taken into account in overall project planning in order to avoid follow-up costs later on caused by rescheduling etc. In addition, the following principles are of prime importance in a **qualification and acceptance program**:

- Implementation and precise application of acknowledged standards
- Comprehensive and complete planning and documentation of all measures and activities
- Communication and cooperation on the basis of clearly defined responsibilities
- Accurate documentation of inspections, identification of all nonconformances or deficits, initiation and follow-up of corrective measures
- Implementation of processes for continuous improvement of the procedures and methods.

A recommended **test sequence** which has proven its worth in practice is shown in Figure 8.3.1.

The reasons for this **test sequence** are the following:

1. The loads on the test object are imposed in the **same sequence** as they will occur during the launch of a rocket: first the mechanical loads, followed by the thermal impact, and concluding with the electromagnetic loads.
2. This test sequence is also based on **cost considerations**. During thermal testing, the costs that arise are significantly higher than those caused by mechanical testing. If repairs are necessary based on the results of the mechanical tests, only these have to be repeated, not the thermal tests, which are much more cost intensive.
3. If all mechanical tests are performed in one sequence, the project benefits from **synergies** that influence both time and cost. The measurement sensors for the testing can be used for all the mechanical tests, making it unnecessary to remove them and newly install them numerous times. This would otherwise have to be done to avoid outgassing the sensors, glue and cables, which would cause contamination during thermal testing. Also, if the thermal balance and thermal vacuum tests are performed in one sequence, the extremely demanding test preparations do not have to be done twice.



Notes:

1 These tests may be combined.

2 In case of larger structures, the random part should be covered by an acoustic test. If random vibration is performed, it should be performed simultaneously with sinusoidal vibration in every axis.

3 These tests may be combined.

4 Performed for completeness. It can also be performed on separate equipment.

5 Equipment for microgravity utilization space vehicle.

6 To be performed for human space vehicle.

Figure 8.3.1: *Typical test sequence (Source [8.3.1]).*

4. For the **vibration test**, it is very helpful to know the exact mass properties, such as mass and center of gravity. Therefore, mass properties should be determined at the beginning. In practice, however, this determination is also often used in order to optimize the whole test campaign, for example to bridge waiting periods related to the availability of test facilities.
5. **Magnetic measurements** should be performed at the end of the test campaign in order to avoid a change in magnetic characteristics caused by any subsequent tests. It could be, for example, that the magnetic characteristics of a test object change after the vibration test because the electrodynamic vibration test facilities create large magnetic fields during testing.

It is quite clear that a well-planned **test campaign** is crucial for the success of the project as it ensures the functionality and thus the success of the mission, adherence to the schedule and to the budget.

8.3.3 Mechanical Tests

Each piece of space equipment is subjected to various **mechanical stresses**, mainly at the beginning of a mission (launch phase). But sometimes also during a mission, for example during trajectory and attitude control maneuvers, separation processes, atmospheric entry and landing, mechanical stress has to be considered.

During launch, apart from static loads due to the launcher's acceleration and the aerodynamic loads, vibrations are generated by the propulsion system and the atmospheric environment (wind gusts, vortices). In addition, the high acoustic noise level during launch (e.g., ground reflection of the sound pressure) and during the transonic flight phase (aerodynamic effects) produces fluctuating pressure fields impinging the surface of structures. Similar mechanical loads occur during atmospheric reentry, augmented by high thermal loads. Moreover, the ignition of propulsion systems, separation and release processes initiated by pyrotechnic devices as well as landing maneuvers result in temporary shock loads.

The suitability of space equipment to withstand these mechanical loads is verified in specific vibration, acoustic, shock and static tests at either the qualification or the acceptance stage.

In addition, for all dynamic processes in the space equipment or during a mission (e.g., attitude control, other flight maneuvers), exact knowledge of the relevant structural parameters like **resonance characteristics** and **mass properties** is extremely relevant. These parameters are determined by using mathematical models. However, before applying the results, the suitability of the mathematical models has to be verified by experiment as well. Modal survey tests and measurement of the mass properties are the experimental tools which provide the required information. These investigations are usually performed on structural models (modal survey) or flight models (mass properties determination).

Before starting a detailed discussion of the mechanical tests, the mechanical loads will be distinguished with respect to their temporal appearance and relevance during a mission.

Quasi-static loads mainly occur during launch, maneuvers and atmospheric reentry. They are evoked by the propulsion systems, aerodynamic loads and inertial loads. They change slowly in terms of time and result in relatively low structural responses.

Transient and shock loads occur when rapid changes of otherwise stationary conditions take place, for instance the thrust build-up or shutdown of propulsion systems, separation processes and landing. These events may lead to very short-term but extremely high structural responses (shock responses).

Quasi-harmonic loads typically occur as a transient harmonic excitation, for example due to instability of the propulsion system, the temporary excitation of structural resonances, or the autoexcitation of dynamic systems caused by back coupling (e.g., the POGO effect). These excitations may last several seconds and may result in significantly high structural responses. In particular for launcher systems, care has to be taken to avoid such occurrences since they may result in dynamic instability beyond control.

Random loads occur during all aerodynamic and propulsion processes. They are introduced into the structure either by body noise or airborne noise. The respective structural responses may be high.

8.3.3.1 Vibration Tests

Purpose

Vibration tests are used to simulate dynamic mechanical loads.

The relevant vibrations for space applications cover a **frequency range** from 1 to 2000 Hz and can be characterized in the following manner:

- Transient or quasi-harmonic in the lower frequency range (1 to 100 Hz)
- Randomly distributed, noise-like, in the higher frequency range (20 to 2000 Hz).

Due to the wide frequency range, vibration loads are critical for the main system structures as well as for subsystems and components.

The objectives of a vibration test are:

- Proof that a specimen can withstand the occurring dynamic loads.
- Determination of component level vibration loads during system-level tests.

In addition, vibration test measurement data can be used up to a certain degree in order to derive the modal properties of a specimen.

Test Specification, Excitation Types, Theory

In correlation with the character of the vibration loads which are relevant for space applications, vibration tests use two different excitation methods:

- Sinusoidal excitation
- Random excitation.

Sinusoidal excitation is characterized by an oscillating motion with a sinusoidal time history. The frequency and amplitude of the motion are altered in accordance with a specified spectrum. The typical time history is shown in Figure 8.3.2.

The **amplitude of the vibration load** is typically specified as acceleration. However, the velocity or the displacement can also be used for this purpose. The correlation between those parameters for a sinusoidal motion is defined as follows:

- Displacement:

$$x = x_0 \sin(\omega t)$$

- Velocity:

$$v = v_0 \cos(\omega t) \text{ with } v_0 = x_0\ \omega$$

- Acceleration:

$$a = -a_0 \sin(\omega t) \text{ with } a_0 = v_0\ \omega = x_0\ \omega^2$$

The progression of the frequency during a test (the sweep rate) is typically conducted exponentially. A frequency change of one octave is carried out at constant time intervals. The parameters required for a sinusoidal test specification are summarized in Table 8.3.1.

The stationary random excitation is characterized by a statistically distributed random motion which fulfills a spectrum in the frequency domain (after averaging).

As for the sinusoidal vibration test, the acceleration is the fundamental parameter. Figure 8.3.3 shows the acceleration time history of a random motion.

The characteristic parameters for a random vibration test can be deduced from the formulas for signals with normal (Gaussian) distributed amplitudes:

- RMS (Root Mean Square) value of the acceleration:

$$a_{\mathrm{RMS}} = \sqrt{\lim_{T\to\infty} \frac{1}{T} \int_0^T a^2(t)\,\mathrm{d}t} \qquad (8.3.1)$$

- Maximum acceleration:

$$a_{\max} = 3\ a_{\mathrm{RMS}} \qquad (8.3.2)$$

Although the maximum amplitude of normally distributed signals is not limited in theory, a limit of

Table 8.3.1: *Sinusoidal test specification.*

Parameter	Example
Frequency range	5–100 Hz
Amplitude spectrum	5–15 Hz: ±10 mm 15–100 Hz: 100 m/s^2
Sweep rate	2 octaves/min

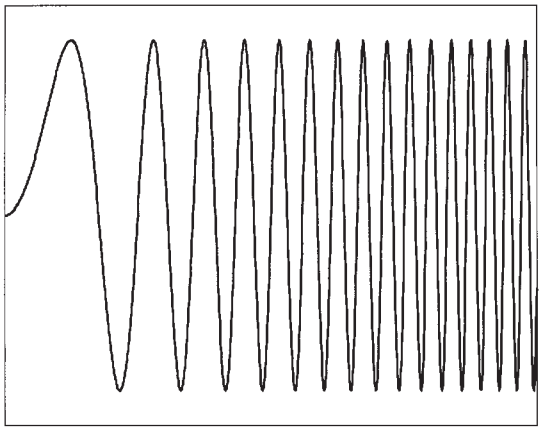


Figure 8.3.2: *Time history of a swept sine test.*

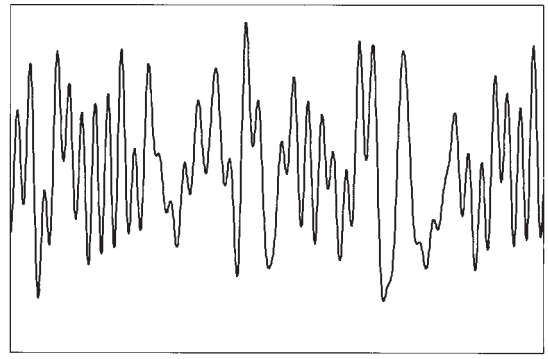


Figure 8.3.3: *Random time history.*

$3a_{RMS}$ (or 3σ) is generally applied in random vibration testing.

A random vibration test spectrum is specified in the frequency domain using the acceleration spectral density (or **power spectral density** (PSD)).

- Power spectral density:

$$PSD(f) = \frac{da_{RMS}^2(f)}{df} \tag{8.3.3}$$

with $PSD(f)$ in $(m/s^2)^2/Hz$ or m^2/s^3.

The required parameters for a random test specification are summarized in Table 8.3.2.

Random vibration can also be simulated during an acoustic noise test.

Vibration test parameters and levels for space applications can be found in the handbooks of the relevant launcher systems. They are derived from measurements or by coupled load analysis.

Test Facilities

The excitation principle for vibration testing is the base excitation (Figure 8.3.4). A rigid support (shaker table) on which the specimen is fixed introduces the vibration motion.

Table 8.3.2: *Random test specification.*

Parameter	Example
Frequency range	20–2000 Hz
PSD spectrum	20–100 Hz: +6 dB/octave 100–1000 Hz: 10 m^2/s^3 1000–2000 Hz: –6 dB/octave
RMS value	120 $(m/s^2)_{RMS}$
Test duration	180 s

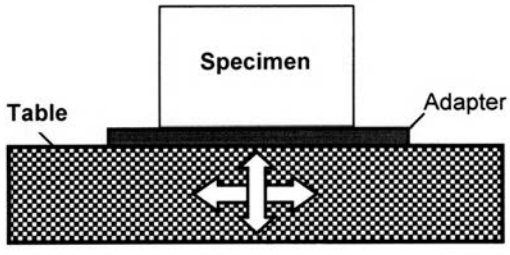


Figure 8.3.4: *Base excitation.*

For testing, hydraulic and electrodynamic **vibration systems** are used. The properties of the two different system types are compared in Table 8.3.3.

Due to the wider frequency range up to 2000 Hz and the higher test accelerations, the electrodynamic vibration systems are those predominantly used. However, hydraulic systems are an important option for large structures with low eigenfrequencies due to their advantages in the lower frequency range and their capability to excite six degrees of freedom.

Test Procedure, Safety Measures

For vibration testing, a specimen is fixed rigidly via a **test adapter** to the table of the vibration system. The vibration system as well as the test adapter should not show any eigenfrequencies in the test frequency range nor generate any cross-excitation axis motions.

The specified vibration loads are determined directly next to the fixation points of the specimen with one or more **control accelerometers**. Whenever several control accelerometers are used, an overall control signal is generated from the single signals by using for example the maximum or average value of the different signals. The basic setup of a vibration test is shown in Figure 8.3.5.

The vibration loads are introduced to the specimen consecutively along its three orthogonal axes.

Due to the high mechanical loads that are applied during vibration testing, the safety of the specimen is of utmost importance. In order to ensure this, several different test runs are executed during a vibration test campaign:

- **Test Adapter Run:** The test is run without the specimen in order to verify the dynamic properties of the facility and the test adapter.

Table 8.3.3: *Comparison of hydraulic and electrodynamic vibration systems.*

Parameter	Hydraulic	Electrodynamic
Frequency range	1–100 Hz	5–2000 Hz
Displacement	±70 mm	±10–20 mm
Acceleration	70 m/s^2	1000 m/s^2
Excitation	6 degrees of freedom (DOF)	1 DOF

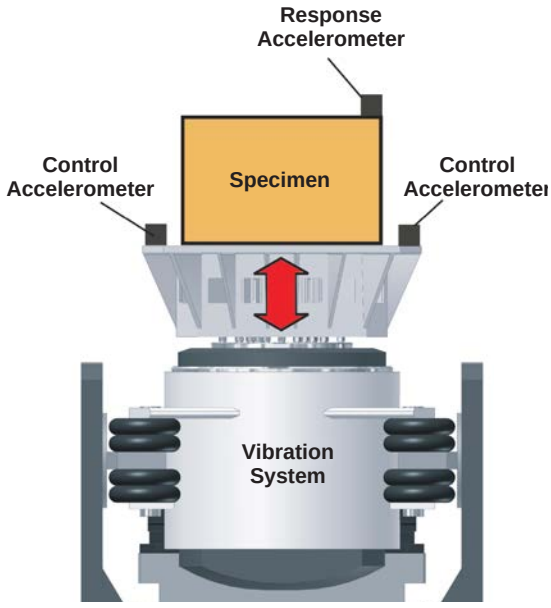


Figure 8.3.5: *Vibration test setup (Source: IABG).*

- **Resonance Search Run:** A low-level sinusoidal test to determine the dynamic properties of the specimen prior to the full load test and in order to check the mechanical integrity of the specimen after the full load test.
- **Intermediate-Level Runs:** Test runs at higher levels in order to check the linearity of the specimen properties.
- **Qualification Runs:** Test runs with the specified loads.

A typical sinusoidal test sequence is, for example:
1. Test adapter run
2. Resonance search run
3. Intermediate-level sine (50%)
4. Qualification-level sine
5. Resonance search run.

The specimen behavior is determined during a vibration test by using standard sensor types for mechanical testing (e.g., accelerometers, strain gauges, load cells, displacement sensors). With respect to the safety of the specimen, limit or abort criteria can be defined for critical measurement locations. If an abort criterion is met, the vibration test is immediately stopped. If limit criteria are met, the vibration input is automatically reduced (so-called **notching**), in order not to exceed the specified limits.

Results
The vibration test results are typically provided as spectra (acceleration versus frequency or PSD versus frequency):

- Control spectrum for the vibration input
- Response spectra on the specimen.

In addition, the acquired signals can be provided as time histories or as frequency response functions.

A vibration test is considered successful if the applied vibration load is in accordance with the specification, if the specimen does not show any mechanical damage after testing, and if the functional test on the specimen after the vibration test is passed without failure.

8.3.3.2 Acoustic Tests

The **acoustic noise test** verifies the suitability of space equipment to withstand the acoustic noise levels anticipated during launch. Often, due to the broadband random excitation, the acoustic noise test is also considered as an acceptance test to reveal hidden failures in **manufacturing**.

The acoustic noise test supplements the vibration test with respect to structural loading in the medium- and high-frequency ranges. The fluctuating noise pressure fields impinge on the test article's surface and cause, in particular on structures with large surface areas but low mass density (e.g., antenna reflectors and solar array wings), significantly high structural responses in the range of up to 200 to 300g_{RMS}. This emphasizes the importance of an acoustic test for such structural components.

Noise fields to be used in tests are implicitly required to be diffuse and homogeneous. That is, the noise impingement is not directional in the statistical sense and the **acoustic spectrum** can be shaped to meet the typical characteristics of an individual launcher. Indeed, for testing, a specific acoustic noise spectrum is used for each launcher type. The acoustic

noise spectrum is expressed in octave or third-octave band **sound pressure levels** (SPL) in decibels (dB). In Table 8.3.4 a typical octave band acoustic noise spectrum is given.

Another parameter of the acoustic noise field is the **overall sound pressure level** (OASPL), which is determined by the octave or third-octave band sound pressure levels as follows:

$$OASPL = 10 \log \sum_i 10 \exp\left(\frac{1}{10} SPL_i\right) \qquad (8.3.4)$$

Typically, the required sound pressure levels for space equipment at the qualification stage amount to 143 to 148 dB_{OASPL}. At the acceptance stage the OASPL is 3 dB lower. Acoustic noise simulation for atmospheric reentry, however, requires sound pressure levels of up to 170 dB_{OASPL}.

Due to the random nature of acoustic noise, the acoustic noise spectrum is always specified along with tolerance bands indicating the lower and upper limits.

An acoustic noise test consists of the following sequence: one test at low level (−8 to −6 dB), one test at **acceptance** level (−3 dB), one test at **qualification** level (0 dB), if required, and the repetition of the low-level test. The structural responses of the two low-level tests (usually acceleration responses, interface forces or strain measurements) shown in **PSD** diagrams are compared and checked for indications of structural change possibly indicating failures. The structural responses obtained from the acceptance or qualification-level test are compared to the design response levels of the individual components of the test article in order to verify the suitability of the design to withstand the acoustic loads. This check is of particular importance, since the acoustic simulation models for acoustic loads depend on statistical and empirical methods whose assumptions have to be verified by test.

Table 8.3.4: *Octave band acoustic noise spectra.*

1/1-Octave band, center frequency [Hz]	SPL qualification [dB]	SPL acceptance [dB]
31.5	132	129
63	137	134
125	140	136
250	143	140
500	139	136
1000	138	135
2000	131	128
OASPL	145.7	142.7

The performance of an acoustic noise test using high acoustic sound pressures requires specific **reverberation chambers**. Reverberation chambers are designed with acoustically reflective walls all around, thus yielding long reverberation times. Long reverberation times are desired since the required energy to produce high acoustic sound pressure levels is minimized. The reverberation chambers for space applications usually provide volumes between 1100 and 1600 m^3. The physical reason for such a large volume is that acoustic energy in reverberant environments can only be established when acoustic modes of the chamber, namely standing waves, can be excited. Accordingly, the largest chamber dimension defines the longest acoustic wavelength and thus the lowest acoustic frequency which can be excited. It is desirable to design a reverberation chamber with a lowest acoustic mode around 20 Hz so that the next acoustic modes can provide a sufficiently high acoustic mode density in the 31.5 Hz and the 63 Hz octave bands. Provided that there is a high acoustic mode density, sufficient homogeneity of the diffuse acoustic noise field can be achieved because it is actually built up by the superposition of numerous acoustic modes.

The **acoustic energy** is mainly produced by using pneumatic transducers. Their working principle is to feed pressurized gas (air or nitrogen) through electrodynamically driven grids which act like a valve. Quick opening and shutting of the grids by using appropriate random drive signals causes pressure fluctuations of the airstreams which, once conducted to the reverberation chamber via exponential horns, provide the required acoustic excitation of the chamber.

In order to control the acoustic noise spectrum, the noise field of the reverberation chamber is measured with a number of microphones (usually 4–12, depending on the application). The averaged acoustic pressure spectrum is compared to the specified spectrum and the driving spectrum for the pneumatic transducers can be adjusted accordingly. In large reverberation

chambers, realistic maximum overall sound pressure levels amount to 153 to 160 dB_{OASPL}. If higher acoustic sound pressures are required, use has to be made of smaller reverberation chambers or progressive wave tubes with much smaller acoustic volumes. In a progressive wave tube, the acoustic noise field generates a progressively grazing noise impingement.

8.3.3.3 Shock

Purpose

Shock tests are used to simulate short-term, mostly transient, mechanical loads. Shock loads are generated during a launcher start by ignition of the propellant systems and especially by the pyrotechnical activation of release mechanisms (e.g., stage separation, fairing separation or satellite release). This type of mechanical load is characterized by high accelerations (up to 100 000 m/s²) and very short impact durations (10–20 ms). Therefore, it is more relevant for electronic, fine mechanical or optical components than for the main system structures.

The objectives of a shock test are:

- Proof that a specimen can withstand the occurring shock loads.
- Determination of component-level shock loads during system-level tests.

Test Specification, Theory

Shock loads can be specified by the corresponding acceleration time history or by using the **shock response spectrum** (SRS) in the frequency domain.

Figure 8.3.6 shows the typical acceleration time history of a pyrotechnically activated shock for space applications.

This transient, random excitation can hardly be described in a time history format. Therefore, the frequency domain description as a shock response spectrum is most commonly used.

A shock response spectrum defines the maximum acceleration responses on an infinite number of dynamic single degree-of-freedom systems caused by a shock event (Figure 8.3.7).

For the relevant damping a value of 5% ($Q = 10$) is generally applied.

The typical **pyroshock spectrum** is characterized by a slope of 9–12 dB/octave followed by a constant level, as shown in Figure 8.3.8.

Shock test parameters and levels for space applications can be found in the handbooks of the relevant launcher systems. They are derived from measurements or by the use of coupled load analysis.

The required parameters for a shock test specification are summarized in Table 8.3.5.

Occasionally, test specifications with impulse-type shock loads (e.g., **half-sine shocks**) can be found. However, these are not suitable for simulating a transient-type shock event.

Test Facilities, Test Procedures

The simulation of shock loads for space applications is performed with either pyrotechnical or mechanical excitation.

For **pyrotechnical excitation** the actual pyromechanisms like the satellite clamp bands are used. In addition, standardized pyroshock setups (so-called **shock generating units** (SHOGUNs)) have been developed. A typical example of this approach is the "SHOGUN-Test" offered by Arianespace for the Ariane 5 shock load verification of satellites.

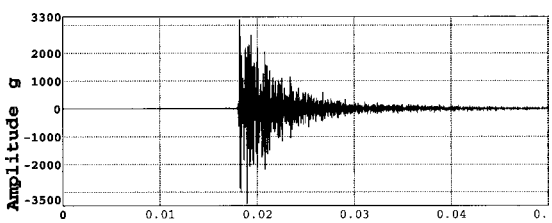


Figure 8.3.6: *Pyroshock time history.*

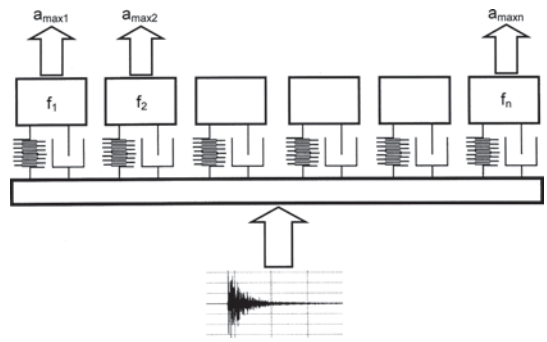


Figure 8.3.7: *SRS definition.*

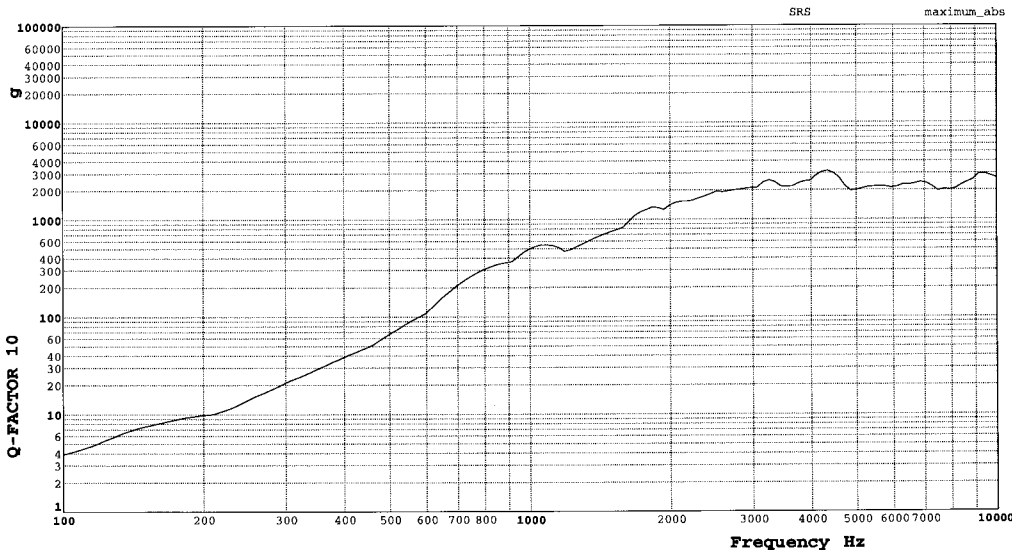


Figure 8.3.8: *Pyroshock SRS spectrum.*

Table 8.3.5: *Shock test specification.*

Parameter	Example
SRS spectrum	100–2000 Hz: +9 dB/octave 2000–10 000 Hz: 30 000 m/s²
Damping	Q = 10 (5%)
Number	Three shocks per axis ($X/Y/Z$)

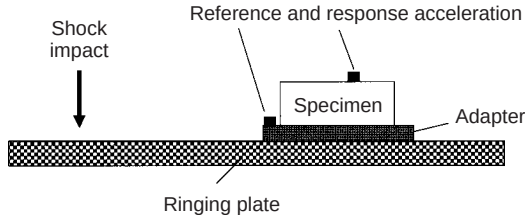


Figure 8.3.9: *Ringing plate setup.*

Pyrotechnical shock simulation is, however, mainly restricted to system-level tests, due to the high corresponding costs. The test procedure is characterized by the required safety measures and the critical coordination between shock generation and data acquisition.

The mechanical simulation of transient shock loads is performed with electrodynamic vibration systems or with the use of so-called **ringing plate test setups**. These less expensive test methods are generally used for component-level tests.

Because vibration systems are limited to maximum accelerations of approximately 7500 m/s² (SRS) and maximum frequencies of 2000 Hz, the ringing plate setups are better suited for pyroshock simulations. A typical test setup is shown in Figure 8.3.9.

The shocks can be generated with different impact methods in this configuration, for example by the impact of a pendulum, by pneumatically induced impacts or by pyrotechnically induced impacts.

Test execution is divided into two parts. During pretests the specimen is replaced by a mass dummy in order to iteratively adjust the facility parameters, so that the specified shock spectrum can be verified at the specimen interface. Thereafter, the specimen is tested with the adjusted shock parameters.

Data Acquisition, Results

During shock tests mostly accelerations are of interest:

- Reference acceleration on the specimen interface
- Response accelerations on the specimen.

The measured data is provided as time histories as well as in SRS format.

A shock test is considered successful if the applied shock load is in accordance with the specification and if the functional test on the specimen after the shock test is passed without failure.

8.3.3.4 Modal Survey

The modal survey test or modal analysis is used to determine the dynamic characteristics of structures within the mission-relevant frequency band in terms of the structure's **modal parameters**: that is, the natural frequencies and the associated modal damping factors, the modes of vibration, and the generalized and effective masses. The relevant frequency range for space applications expands up to 60 Hz or 150 Hz, depending on the application. The modal parameters are regularly used to verify the structural models and to update them, if required.

Modal tests are performed on space equipment if verified mathematical models are required for coupled load analysis for the launch configuration, for launcher guidance or when very specific dynamic conditions have to be met (e.g., if the lowest eigenfrequencies with high effective mass of a payload have to be higher than a certain defined frequency). Representative example projects for which modal survey tests had to be performed comprise all major components of the Ariane 5 launcher, the Space Shuttle and all Space Shuttle payloads, as well as payloads with high effective mass.

For a modal survey, the **equation of motion** for a discrete mass system is considered:

$$\boldsymbol{M}\,\ddot{\boldsymbol{x}} + \boldsymbol{F}_{\mathrm{D}} + \boldsymbol{K}\,\boldsymbol{x} = \boldsymbol{p} \tag{8.3.5}$$

$\boldsymbol{M}$ and $\boldsymbol{K}$ represent the mass and the stiffness matrix, respectively, $\boldsymbol{x}$ the vector of the displacement degrees of freedom (during the modal test practically the measured structural responses) and $\boldsymbol{p}$ the vector of the external excitation. The damping forces $\boldsymbol{F}_{\mathrm{D}}$ are described either with the simple linear viscous approach $\boldsymbol{F}_{\mathrm{D}} = \boldsymbol{D}\dot{\boldsymbol{x}}$ or by using the approach for structural damping $\boldsymbol{F}_{\mathrm{D}} = i\boldsymbol{D}\boldsymbol{x}$. Both damping models describe only approximately the real damping behavior since real damping processes do not follow linear models in general and often several physical damping mechanisms act in parallel. Therefore, it is sufficient to assume the simplest case of structural damping in the following equations.

Basically two different techniques exist to determine the modal parameters, namely the phase resonance and the phase separation technique.

Phase Resonance Technique: At selected points several exciters are attached to the structure so that the individual eigenmodes can be excited directly and isolated from all other modes. Then the modal parameters of such an isolated mode are directly accessible by measurement. This technique, also called the **ground vibration test** (GVT), uses harmonic excitation to vibrate the structure under test while keeping the phase of all excitation forces real (i.e., the phase angle is 0° or 180°). The harmonic excitation frequency and the distribution of the forces (in amplitude and phase or antiphase, respectively) are adjusted such that the structure's response is in resonance; that is, the structural responses are in quadrature with the excitation forces. For a real excitation force pattern this means that the real parts of the structural responses are zero and the imaginary parts of the responses are at a (local) maximum.

Under this tuned condition, the equation of motion can be resolved into an imaginary and a real part:

$$\left[-\omega^2\boldsymbol{M} + \boldsymbol{K}\right]\cdot\boldsymbol{x}^{\mathrm{re}} - \boldsymbol{D}\cdot\boldsymbol{x}^{\mathrm{im}} = \boldsymbol{p} \tag{8.3.6}$$

$$\left[-\omega^2\boldsymbol{M} + \boldsymbol{K}\right]\cdot\boldsymbol{x}^{\mathrm{im}} + \boldsymbol{D}\cdot\boldsymbol{x}^{\mathrm{re}} = 0$$

The result is the fundamental physical condition that the excitation forces are balancing the structure's damping forces and the isolated eigenmode is vibrating in its undamped frequency:

$$-\boldsymbol{D}\cdot\ \boldsymbol{x}^{\mathrm{im}} = \boldsymbol{p};\quad \boldsymbol{x}^{\mathrm{re}} = 0 \tag{8.3.7}$$

$$\left[-\omega^2\ \boldsymbol{M} + \boldsymbol{K}\right]\cdot\boldsymbol{x}^{\mathrm{im}} = 0$$

Hence, in the tuned condition, the eigenfrequency ω_r and the respective normal mode $\Psi_r = \boldsymbol{x}(\omega_r)$ are directly accessible to the measurement. The structural damping parameter g_r is determined by analyzing response functions around resonance measured by varying the excitation frequency while maintaining the tuned force pattern. Finally, the modal mass m_r is determined by the energy input, that is the scalar product of excitation forces and the acceleration responses at the excitation points, by means of a simple algebraic equation and using a suitable normalization factor n_r:

$$m_r = \frac{1}{g_r \; n_r} \cdot \boldsymbol{x}_r^T \; \boldsymbol{p}_r . \qquad (8.3.8)$$

In addition, for experimentally determining the effective masses the interface forces at the restrained degrees of freedom also have to be measured. More detailed background on the mathematical and experimental concepts may be found in the literature [8.3.6].

Phase Separation Technique: Any dynamic structural response spectrum of a linear system may be reconstructed by a linear combination of all modal responses based on the modal parameters and the eigenmodes Ψ_r:

$$x(\omega) = \sum_{r=1}^{N} \frac{\boldsymbol{\psi}_r^T \boldsymbol{p}}{m_r} \frac{\boldsymbol{\psi}_r}{\left(\omega_r^2 - \omega^2\right) + i \; \eta_r \; \omega_r^2} ; \qquad (8.3.9)$$

Vice versa, in a least squares approach, any measured structural response spectrum may be decomposed into single linear damped oscillators by mathematical appropriation. In this manner, the modal parameters and the mode shapes may be extracted from measured response spectra. This concept is called **modal analysis** and involves the mathematical extraction of modal parameters from already measured structural responses. On the other hand, the term **modal test** also covers the specific experimental effort to provide the required structural response spectra in a dedicated test setup. The respective detailed procedures and the subsequent application of phase separation techniques are manifold but well documented in the literature; reference is made to relevant monographs like [8.3.6] or other technical literature.

With this background, the most suitable procedures to determine modal parameters for space equipment can now be discussed.

Phase separation techniques are very flexible in their application since the input may be response spectra, frequency response functions, or transmissibility functions which may have been obtained from other mechanical tests, in particular vibration tests. As a consequence, this approach is quite popular since it is comparatively less expensive and less time consuming than tests based on phase resonance techniques. However, it should be noted that this flexibility entails extracted modal parameters of varying quality; in particular, the modal mass cannot be determined when the excitation force is not known (e.g., when using transmissibility functions) for a model analysis. As a consequence, selecting the most appropriate procedure to determine the modal parameters cannot just be a matter of inherent cost and time aspects; the accuracy requirements for the test results still play an important role and have to be considered in terms of the purpose that the modal results have to fulfill. The relation between test effort, quality of results and test objectives is depicted in Figure 8.3.10.

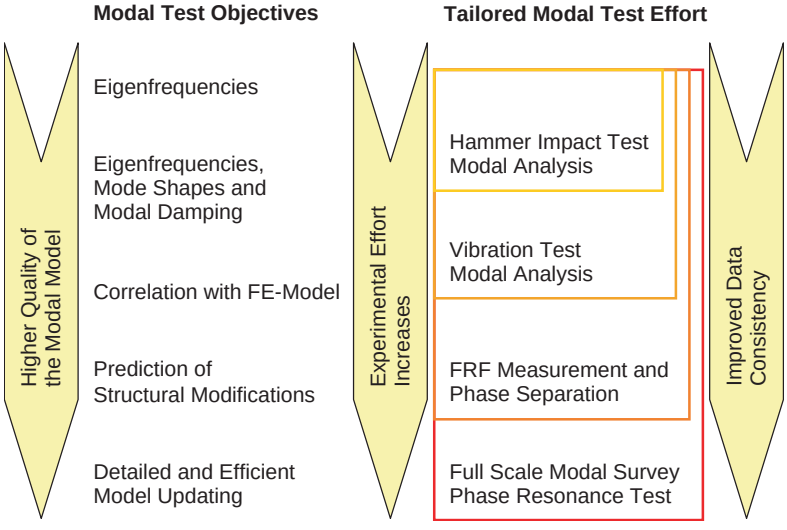


Figure 8.3.10: *Relationship between quality of modal test results and the required effort.*

Provided that for a model verification only eigenfrequencies and mode shapes are required and the modal damping is to be determined, simple modal test approaches to measure structural responses by applying hammer impact tests, or structural responses obtained from a vibration test, may be sufficient for modal analysis. However, for exact model verification and updating a mathematical dynamic model, a dedicated modal survey test is required. The classical ground vibration test using phase resonance techniques will always provide more accurate results than multipoint random frequency response measurements with a subsequent modal analysis.

The physical and procedural reasons which speak in favor of using the high-quality standard of ground vibration test results are as follows:

- During a ground vibration test all measurements and conditions are accessible in a clear manner and the modal parameters can be determined directly (excitation forces, behavior of interfaces and restraint conditions, check of the structure for linear behavior, direct observation, or even modeling the correction of nonlinear effects in the measurements).
- The modal parameters are determined in a steady-state vibration condition.
- Quality checks on the modal parameters may be performed, and immediate correlation with the analytical test prediction is feasible at any time during the test process. This enables pinpointing of any inconsistency and doubt which may arise during the experimental investigation.
- The accuracy of the main modes can be determined to high precision on the basis of a comprehensive and consistent data basis.

Compared to this list, **modal analysis of vibration test data** has the following disadvantages:

- Some of the test parameters are not exactly, or only with difficulty, or not at all accessible to a modal analysis. This involves, for instance, the speed of the swept sine excitation, which does not allow for a complete build-up of a steady-state modal vibration, or for potential interaction of the vibration table and the closed loop vibration control system with the dynamic characteristics of the test article. These effects can only partially be accounted for by mathematical or empirical corrections. In certain circumstances, namely a test article with high mass, even the mathematical model of the vibration table has to be considered when comparing test results and model prediction.
- The excitation forces acting between the vibration table and the test article are not accessible to measurement, unless a dedicated interface force measurement device is used. If these forces are not measured, the modal analysis cannot provide the modal mass or the effective masses based on the experimental data. These modal parameters can only be estimated by making reference to the analytical mass matrix of the test article.
- Even if the interface forces and moments between the vibration table and test article are measured, the resulting frequency response functions describe the dynamic behavior of the vibrating system in free–free boundary conditions. The mathematical transformation into a restrained system requires additional assumptions.
- The methodically proper resolution of closely spaced modes is excluded since the required orthogonal excitation patterns are missing.
- The estimation of the modal parameter's accuracy is quite difficult and may even lead to unreasonably high and not useful error bounds.

Nevertheless, the modal analysis of vibration test data has emerged as a practicable and inexpensive tool to provide an overview of the modal characteristics, taking into account that the resulting accuracy is not as high as that obtained from a dedicated modal survey test. Accordingly, the improvement of this modal analysis approach is still a subject for scientific research.

Modal test results are correlated with the respective analytical models by using the orthogonality relations of eigenmodes:

$$\boldsymbol{A}_{rs} = \boldsymbol{\psi}_r^T \, \boldsymbol{M} \, \boldsymbol{\psi}_s = \begin{cases} 0 & \text{for } s \neq 1 \\ m_r \text{ (real)} & \text{for } s = 1 \end{cases} \qquad (8.3.10)$$

The close interaction of test results and analytical results requires the dedicated analytical preparation of a modal survey test. This comprises the identification of target modes, the derivation of a measurement plan

enabling satisfactory observation and identification of these eigenmodes, and the provision of a condensed mass matrix and of predicted mode shapes allowing correlation of the test results with the mathematical model. In this respect, many modal survey tests have proved that thorough test preparation makes a decisive contribution to achieving the test objectives.

8.3.3.5 Mass Properties

The determination of **mass properties** serves as the exact experimental verification of all dynamic properties of the rigid body, like mass, the location of the center of gravity (CoG), the moments of inertia (MoI), the products of inertia (PoI) and the orientation of the **principal axes**. These properties are used to confirm the mass budget calculations. This is a quite complicated process due to the many components of a spacecraft (also considering attachments, interfaces, bolts, cabling). These calculations demand much effort and care, thus necessitating experimental verification of the as-built configuration.

The motion of a rigid body of mass m is described by Newton's fundamental law in the form $\boldsymbol{M}\ddot{\boldsymbol{x}} = \boldsymbol{p}$, which is composed of the vector of the rigid body translations and rotations $\boldsymbol{x}^T = [x\ y\ z\ \varphi_x\ \varphi_y\ \varphi_z]$ and the vector of the external forces and moments $\boldsymbol{p}^T = [F_x\ F_y\ F_z\ M_x\ M_y\ M_z]$ referenced to any arbitrary point. The respective mass matrix consists of 10 independent elements:

$$\boldsymbol{M} = \begin{bmatrix} m & 0 & 0 & 0 & mz_c & -my_c \\ 0 & m & 0 & -mz_c & 0 & mx_c \\ 0 & 0 & m & my_c & -mx_c & 0 \\ 0 & -mz_c & my_c & J_{xx} & J_{xy} & J_{xz} \\ my_c & 0 & -mx_c & J_{xy} & J_{yy} & J_{yz} \\ -my_c & mx_c & 0 & J_{xz} & J_{yz} & J_{zz} \end{bmatrix}.$$

The determination of mass properties is usually performed on flight models, which provides the most accurate results for launch and enables exact prediction of the spacecraft's reaction to the attitude control system. Accordingly, the determination of the mass properties has to fulfill the highest accuracy requirements. When determining the mass, the accuracy is usually better than 0.5%, better than 1 mm for the center of gravity, and better than 2–5% for the moments of inertia.

Direct determination of the mass properties using Newton's fundamental law of motion by means of force and acceleration measurements is too inaccurate for this purpose. Accordingly, highly precise weight scales, center of gravity scales and torsion pendulum machines have to be used.

The practical problem, however, is that the center of gravity or the principal axes of the test article must be moved into the measurement axis of the mass property scales. Figure 8.3.11 shows the measurement setup for the determination of the center of gravity along the vertical axis of the spacecraft, and the moments of inertia around its lateral axes.

In order to facilitate handling and to reduce work for the mechanical setup, new generalized **torsion pendulum machines** have recently been introduced. The working principle of the generalized torsion pendulum is shown in Figure 8.3.12. This machine allows the test article to be moved along the vertical bar of the machine and the test article to be rotated around its vertical and lateral axes.

The generalized torsion pendulum machine also incorporates center of gravity scales. Accordingly, the measurement procedure is as follows:

1. Determination of the mass on separate weight scales.



Figure 8.3.11: *Determination of the center of gravity along the sic's vertical axis and of the moments of inertia around its lateral axes (Source: IABG).*

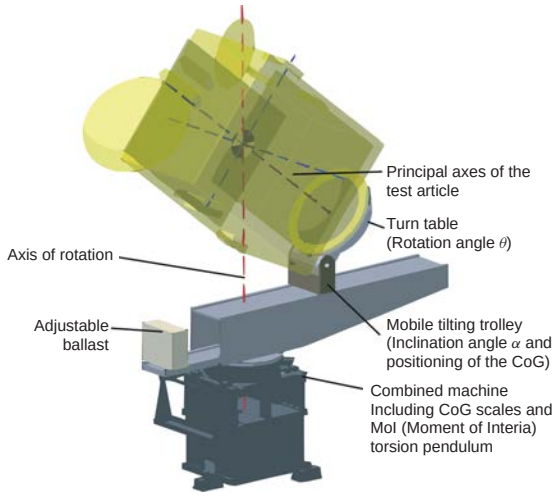


Figure 8.3.12: *Working principle of the generalized torsion pendulum (Source: IABG).*

2. Determination of the lateral coordinates of the center of gravity and of the vertical axis while the test article is placed in a vertical orientation on the generalized torsion pendulum.
3. Localization of the center of gravity along the vertical axis by using the torsion pendulum. The test article's vertical axis is inclined and moved step by step in the horizontal direction on the machine's bar. If the center of gravity coincides with the rotation axis of the torsion pendulum, then Steiner's part of the product of inertia vanishes and the frequency of the torsion pendulum is at maximum.
4. Now, the center of gravity is always kept in the axis of the torsion pendulum and by inclining the test article's vertical axis, or by rotating around the test article's vertical axis, the products of inertia can be measured around various axes of rotation. Since the geometrical orientation of the rotation axis with respect to the axes of the test article is known, the coefficients *A*–*F* can be determined, and by a series of measurements of the product of inertia Θ, an overdetermined system of equations can be established to evaluate all **moments and products of inertia J_{xx} to J_{xz}** in the least squares sense:

$$\Theta = AJ_{xx} + BJ_{yy} + CJ_{zz} + DJ_{xy} + EJ_{xz} + FJ_{yz}$$

8.3.4 Space Simulation Tests

To verify the characteristics and behavior of satellites in a space environment, they undergo extensive **thermal/vacuum tests** during the verification process**.** Unlike the mechanical tests which mainly test the behavior during the launch phase, thermal vacuum tests concentrate on the qualification of the satellite for the operational phase later in space.

The **space conditions** simulated during thermal vacuum tests are the pressure level in high-altitude orbits, the cold space background and various thermal sources (solar radiation, albedo and Earth radiation) along with the influences they exert on the **thermal balance** of a satellite. Due to space vacuum, heat exchange is mainly determined by radiation, while solar radiation and the cold space background lead to high thermal radiation gradients with correspondingly high temperature gradients on the satellite.

A satellite consists of a multitude of materials, components and subsystems, all of which have to be kept in their specified operational temperature range to ensure proper functioning (e.g., electronic subsystems) or to avoid material damage and shape distortions on structural parts (e.g., antenna reflectors).

To ensure compliance with the temperature ranges in all mission phases, a wide range of passive and active temperature control mechanisms may be utilized, starting with selected surfaces (black, white, metallic, etc.) with well-defined absorption and emission properties (α/ε values) and continuing with multilayer insulation (MLI) foils, heat straps, heat pipes and temperature-controlled electric heaters. All the temperature control measures taken together are commonly called the **thermal control subsystem** (TCS), although it cannot be treated as an independent subsystem like a satellite bus or an antenna reflector.

In the past TCS configuration has mainly been an experimental process based on test data. Today, most of the configuration process can be performed using the results of numerical calculations. With the numerical capacity of today's computers, extensive numerical calculations with thousands of nodes are possible; the main challenge is in the exact description

of all relevant thermal and physical properties (α/ε values, thermal masses, etc.) and their interactions.

8.3.4.1 Thermal Balance Tests

An important task of thermal vacuum tests, especially **thermal balance** (TB) tests, is to provide measurement data for specific thermal load cases. This data allows verification and optimization of the satellite's numerical thermal model. Generally, it is not necessary to verify all the possible load cases of a mission by testing; evaluation of some specific (extreme) cases is normally sufficient. It can be assumed that other cases can be correctly evaluated if the (adapted) thermal model can correctly verify the measured cases.

The thermal balance test is normally performed with a cold chamber shroud (cooled with liquid nitrogen) at temperatures below −180 °C to simulate the space background. Each mission phase is then represented by a corresponding solar radiation flux level or heated thermal plates (albedo and Earth radiation simulation). After stabilization the resulting temperature equilibrium on the satellite is measured and recorded.

8.3.4.2 Bakeout Test

Testing space-related components (solar panels, electronic units, antennas, etc.) is another important task of thermal vacuum tests. The main aspect is qualification for later use in space. A so-called **bakeout test** is performed for every important part of a satellite that goes into space. During this test, it is exposed to vacuum in order to force **outgassing** in the operational temperature range of the component.

In vacuum the test specimen is heated up to a temperature level above the operational temperature. Depending on the mission parameter and type of component, a typical temperature range of about +80 °C to +150 °C is used. The test specimen is exposed to such an environment for at least 24 hours, and sometimes significantly longer.

On the ground, the bakeout test ensures the outgassing of most of the volatile components (hydrocarbons, silicones, esters, water, etc.) coming from the production process in order to prevent the satellite from becoming contaminated. The satellite speed in orbit is in the same range as the molecular speed of outgassed material at a total pressure level of $< 10^{-5}$ mbar (the satellite may even be faster), which is the reason why the material does not detach from the satellite. In fact, the components form a cloud around the satellite and condense on the satellite's cold surfaces. This is a huge problem for the optical subsystems or solar cells, for instance, which may become coated and fail.

8.3.4.3 Thermal Vacuum Test

Bakeout tests are often combined with **thermal vacuum cycling** (TV or TVC) tests. Here either space components or the whole satellite are exposed to several **temperature cycles** under vacuum conditions. The number of cycles, as well as the hot and cold temperature extremes, have to be defined with respect to the mission parameters. Typically, the lower temperature is in the range of −100 to −180 °C, whereas the upper temperature of the cycles is normally in the range of +80 to +130 °C. Nevertheless, specific missions may require hot phases of up to +200 °C, which can be realized in specially designed thermal vacuum chambers (e.g., the 2 m thermal vacuum chamber of IABG mbH in Ottobrunn, Germany).

The test specimen is brought to the desired temperature by heat exchange, either by radiation with the thermal shroud of the test chamber or directly by a thermal plate on which the test specimen is mounted.

Particularly large test items or those with thermal properties that do not allow good temperature adjustment by radiation require a long test duration during TVC tests and are therefore very expensive. Sometimes only parts of the test item – like metal inserts inside a carbon fiber antenna – are responsible for extending the test time.

The maximum or minimum temperatures for single parts of a test item may be different from those for the main part of the test item. For example, a waveguide in an antenna tower assembly may need to be tested at a higher temperature level than what is required for the antenna reflector itself. It is also possible that during the cold phase specific antenna components may not be allowed get as cold as the antenna itself.

In both cases additional heating elements have to be provided in the test setup, allowing specific areas to have temperature levels differing from what is achieved with the chamber's thermal shroud. Either infrared radiators or directly applied electric heaters are used for the purpose. The development of a thermal vacuum test concept meeting the customer's requirements is one of the main tasks of a test center.

8.3.4.4 Thermal Cycling Test

If thermal cycling can be performed without vacuum, the test duration – and the costs – can be reduced significantly. During a **thermal cycling** (TC) test or an **ambient pressure thermal cycling** (APTC) test in an APTC chamber, a dry nitrogen gas atmosphere is heated and cooled under atmospheric pressure to quickly change the temperature of the test item using direct convection. Because the medium circulates at a low flow speed inside the chamber (approx. 1–3 m/s), high temperature gradients from about 5 to above 30 °C/min are possible. The duration of the test is significantly influenced by the mass of the test article.

Subjecting parts of the test article to their own individual temperatures is, however, not possible in an APTC chamber, something which has to be taken into account when developing the test concept.

An APTC test is normally combined with a bake-out test or a thermal vacuum cycle so that the test item is exposed to vacuum at least for a short duration in order to accelerate the outgassing process.

8.3.4.5 OSTC Test

At the end of their production, solar panels undergo a special **on-stage thermal cycling** (OSTC) test.

During this test a solar panel is exposed to **temperature cycling stress** in order to simulate the conditions of its subsequent operation in near-Earth orbit. In orbit the solar panel's cell side always faces the Sun, while the back faces the cold space background. This results in high temperature gradients through the thin panel structure. Additionally, in Earth orbit the solar radiation is periodically interrupted when the spacecraft enters the umbra.

To simulate this **load profile** in a test and thus qualify a solar array, the thermal vacuum chamber is first cooled to a temperature below −180 °C using liquid nitrogen. A strong infrared lamp unit facing the cell side of the solar panel periodically heats that side to the required qualification temperature. A radiation intensity of 8 kW/m^2 causes heating of the solar panel from −170 to +130 °C in less than 30 min. After the infrared source is switched off, the panel cools down again, which may take up to 7 h (cooling to −175 °C), depending on the required minimum temperature (Figure 8.3.13).

The cycles are combined with preceding temperature regulating phases which heat up the test specimen in specific stages, first under ambient pressure conditions and then under vacuum.

Such a test sequence reveals possible fabrication faults such as chipped solar cells, structural cracks, unexpected color changes, or detachments of the panel substrate.

8.3.4.6 Thermal Distortion Test

Heavy geometric demands are made on antenna reflectors in space. These reflectors are shaped according to the transmission region on the ground in order to follow regional borders (countries, continents) as closely as possible. The reflector shape must not vary under thermal loads or after outgassing. The limit for dimensional stability is in the range of a few microns. The stability is measured optically in a thermal vacuum chamber by using high-resolution calibrated photographic cameras (digital photogrammetry, videogrammetry). These tests are also called **thermal distortion** (TD) tests.

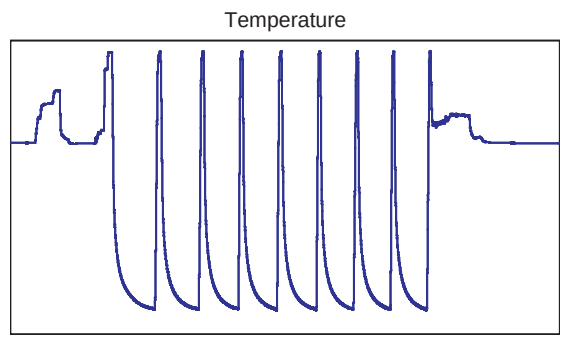


Figure 8.3.13: *Temperature chart of a typical thermal vacuum cycling test.*

A grid of optically reflecting measurement points (retroreflecting targets) is attached to the reflector surface whose 3D geometry has to be evaluated. Specific code targets are used to evaluate the absolute orientation of all target points using digital image processing in order to automate the evaluation process (Figures 8.3.14 and 8.3.15).

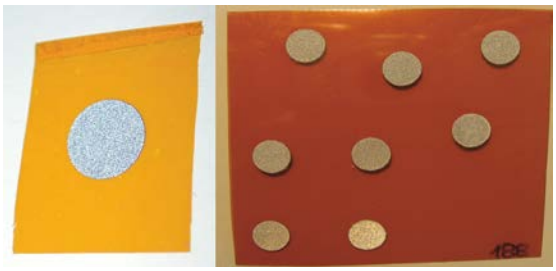


Figure 8.3.14: *Target on a carrier foil, code target for the thermal vacuum chamber (Source: GDV).*



Figure 8.3.15: *Installation of a reflector prepared with targets in the thermal vacuum chamber (Source: IABG).*

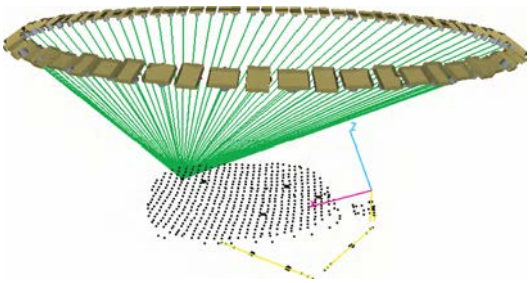


Figure 8.3.16: *Measurement configuration for an antenna reflector (72 camera locations/measurement pictures) (Source: GDV).*

Multiple pictures from different angles are taken during the test program with a digital measurement camera (with a ring flash). Using 3D triangulation the actual position of the targets on the reflector can be calculated ("bundle adjustment"). High measurement accuracy is achieved by precise picture measurement (RMS < 0.3 µm), ideal ray cut angles and high overdetermination of the measurements (Figure 8.3.16).

During a **thermoelastic distortion test** the test specimen surface is measured under the following environmental conditions:

1. Ambient pressure and temperature
2. Vacuum, ambient temperature
3. Vacuum, low qualification temperature
4. Vacuum, high qualification temperature
5. Vacuum, ambient temperature
6. Ambient pressure and temperature.

By comparing the measured distortion at ambient temperature before and after the test (measurements

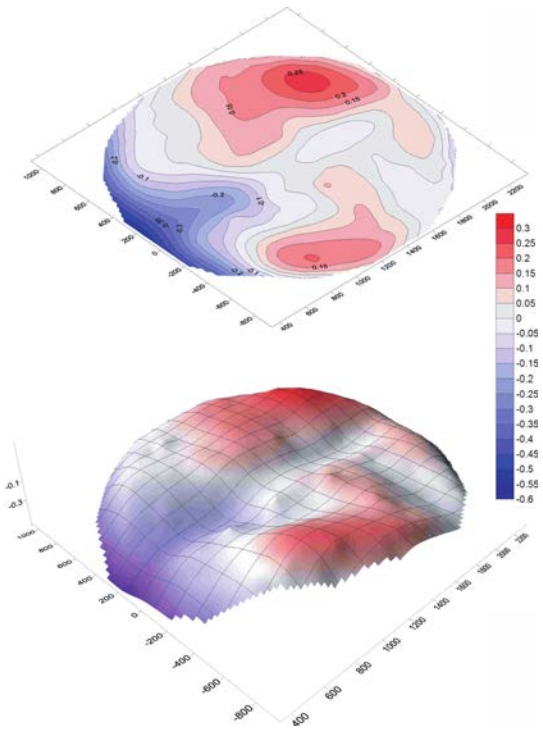


Figure 8.3.17: *Example of a graphical analysis (deformation analysis) (Source: GDV).*

1 and 6) conclusions can be drawn from the remaining **plastic deformation** caused by outgassing. The distortion level at high and low qualification temperatures (measurements 3 and 4) indicates the range of the test specimen's **thermoelastic distortion** (Figure 8.3.17).

8.3.4.7 Test Facilities

Complex **space simulation test facilities** or simple **thermal vacuum test facilities** may be designed and built very differently in detail. However, all facilities need to fulfill the basic requirements for the test of space equipment, namely the simulation of:

- Vacuum
- Cold background radiation
- Solar radiation
- Earth radiation, albedo radiation, etc.
- High temperatures.

A typical facility is shown in Figure 8.3.18. The numbers in the following discussion relate to this diagram.

The main components of a thermal vacuum test chamber are a stainless steel vessel with a connected vacuum system and a black-painted thermal shroud inside ⑤, driven by a thermal system. With this equipment, the test specimen can be exposed to vacuum and heated or cooled by the thermal wall. Every thermal vacuum test chamber meets these basic prerequisites. For space simulation tests a **solar simulator** is additionally required. A projector with an integrated lens system ② and a collimation mirror ③ directs the light from high-pressure xenon lamps ① first through a quartz window into the vacuum chamber and then into the test area ④ in order to simulate the natural collimation and intensity of sunlight. A **motion simulator** enables orientation of the satellite with respect to the artificial solar beam.

The **temperature** of the test specimen is the most important measurement in space simulation. Therefore, up to 1000 thermocouples are attached to the specimen and their values scanned and recorded at intervals of 10 s to 1 min.

The Vacuum System

Requirements related to the vacuum system are derived from the demand that heat transfer due to convection by any gas components present has to be negligible compared to heat transfer due to radiation. It is assumed that the mean free path of the gas molecules is long with respect to the distances between surfaces involved in the heat exchange. This requirement is normally fulfilled with a chamber pressure $<1.3\times10^{-5}$ mbar. This pressure level (or simply "10^{-5} mbar") is therefore typically defined as the requisite criterion for thermal vacuum tests.

The design of a vacuum system is performed for a working pressure of $<10^{-6}$ mbar, also taking into account the outgassing rates of the facility and the test items.

Cryopumps and oil-free turbomolecular pumps are mainly used for the vacuum supply. This type of pump ensures that no contamination of the chamber or the test item with pump oil can occur. Cryopumps have no movable parts. The pumping effect is achieved only by freezing the gas components on cold surfaces. Furthermore, they provide a high pumping capacity in high vacuum, which makes them particularly suitable for large test facilities. Six cryopumps are installed at the space simulation facility of IABG. With individual valve gates they can be separated from the test chamber for recovery during test operation. For smaller chambers often only a turbomolecular pump is sufficient. Since high vacuum pumps can only be used below a pressure level of 10^{-1} mbar, the vacuum vessel is initially evacuated with mechanical pumps (e.g., Roots pumps).

The Thermal System

Liquid nitrogen is generally used as a working medium in thermal systems to cool the black-painted thermal shrouds down to a temperature level of 80–100 K. The simulation errors resulting from the higher background temperature compared to the exact background temperature of about 2.7 K are usually small and can be taken into account in numerical simulations.

A fixed tank facility with pressure regulation is used for the liquid nitrogen supply, which allows an adjustable pressure level of 1–5 bar.

For heating up the chamber after testing and especially for thermal cycling, a blower and heater system are necessary to heat the nitrogen flow to the required test temperature. For this procedure, the gas stream is cycled in a closed loop and conditioned with electric heaters and heat exchangers. The performance of the thermal system is mainly defined by the achievable heating and cooling gradients on the one hand and

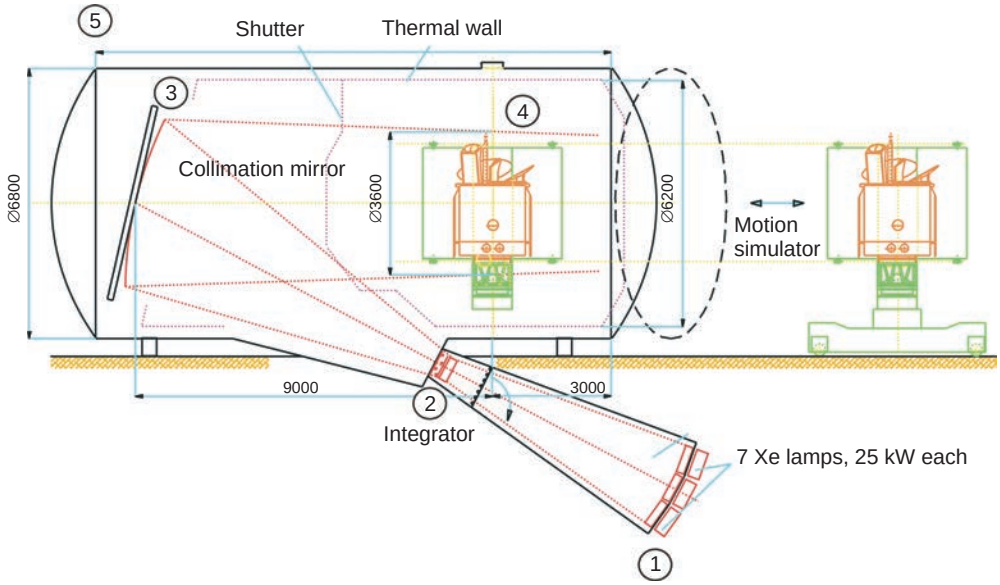


Figure 8.3.18: *Sketch of IABG's space simulation test facility (Source: IABG).*

by the reachable temperature uniformity on the thermal shroud on the other hand. Particular attention is paid to the segments of the thermal shroud. These surfaces are supposed to simulate the cold space background and are therefore painted with a space-proven black paint with high absorption (e.g., $\alpha = 0.95$) and high emissivity ($\varepsilon > 0.9$). The layout of the nitrogen pipes is also an important performance criterion for the thermal system. On the one hand, all parts of the shroud segments have to be uniformly brought to the desired temperature using both liquid and gaseous nitrogen, but on the other hand no gas reservoirs should remain in the shroud while in liquid nitrogen mode since this would cause high temperature gradients inside the shroud.

Solar Simulation

The solar intensity in near-Earth orbit is 1371 W/m^2 (solar constant, *SC*) with a collimation of about 32 arc minutes. **Solar simulators** in the major European test centers (see, e.g., Figure 8.3.19) are able to generate a light intensity up to approx. 1900 W/m^2 (1.4*SC*) in a test plane with a collimation of ±2°, which is a value sufficient to prevent simulation errors caused by shadows. Further typical requirements for solar simulation are of course uniform distribution across the test plane (e.g., < 4%), but also in the defined testing volume (e.g., < 5%); the temporal stability of the light source; and the spectral distribution of the light itself. The last one is mainly defined by the light bulbs used. The high-pressure xenon bulbs with a power consumption of 20–25 kW per bulb that are usually used reproduce the **solar spectrum** in a sufficient way. Only the **xenon peaks** visible in the xenon spectra cannot be found in the natural solar spectrum. For solar simulation it is usually sufficient to reproduce the total intensity. However, the xenon peaks have an effect on the power drain of the solar

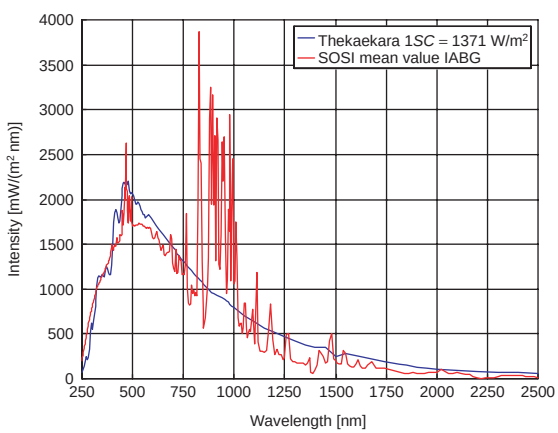


Figure 8.3.19: *Comparison of the Thekaekara spectrum with the sprectral distribution of the solar simulator of IABG (Source: IABG).*

cells, for example, due to spectral effects of the cells: a solar cell illuminated by xenon light produces about 22–26% more electrical power than does natural solar light at the same total intensity.

Three main **optical components** are installed in solar simulators of large test facilities: a light source, an integrator consisting of many lenses, and a collimation mirror. For the illumination of the test plane (diameters of 3–6 m are installed in Europe), between 7 and 20 **high-pressure xenon bulbs** are needed to reach the required intensity. Spare bulbs have to be available in case of failure. Each bulb has its own power supply and cooling system. Cooling is also a main issue for the integrator. In the focus the light of the xenon lamps is guided through a lens system that would heat up strongly without cooling. Here the challenge is to design a highly transparent lens system with cooling loops for every single lens involved.

The big **collimation mirror** is the most eye-catching component of the solar simulator. Consisting of a large number of individually adjusted and cooled mirror segments, it reflects the light coming from the integrator into the test plane. Each individual segment is adjusted according to a complex procedure so that the defined uniformity and geometry can be achieved on the test plane.

Having the collimation mirror at the right temperature has another important function: if not protected by a shutter which has been shroud cooled with nitrogen, the mirror would also face the satellite when solar simulation is switched off. It would thereby represent a totally different heat source for the satellite than do the black surrounding thermal shrouds. This has to be taken into account, for example, in numerical simulations. In order to protect the mirror itself from contamination by condensation of outgassed components, it is always kept at a higher temperature than the shroud and the test item. Cooling and heating of the individual mirror segments are mostly performed by a separate nitrogen circuit of the thermal system.

The Motion Simulator

The motion simulator represents the interface between the satellite and the test facility. Its main task is to position the satellite, which is connected to the spin drive via a thermal test adapter, in the required orientation with respect to the direction of the sunlight or the gravity vector. The swivel angle can be adjusted by ±180°, the spin by 360°. To simulate the **spin stabilization** of a satellite, the spin axis can be continuously rotated at several rotations per minute (up to 10 or 20 rpm). All measurement signals (e.g., 500 temperature signals) and supply lines from and to the satellite have to be connected to corresponding slip-ring units. In addition to a supported satellite mass of up to 5 tonnes, the motion simulator moves its own arm, weighing 20 tonnes, with a positioning accuracy of a few hundredths of a degree. At the same time the self-deformation of the motion simulator due to gravity has to be kept to a minimum. Therefore, it has to be designed so it is stiff enough. High positioning accuracy is also required to test the function of **heat pipes** and the deployment of antennas at the satellite. Since those components are designed for zero-gravity operation, accurate orientation with respect to the gravity vector is necessary.

The motion simulator should not induce any thermal effects on the test specimen. It is therefore covered with a thermal shroud and has the same temperature as the chamber shroud. Since all actuators and control components of the motion simulator have to be sufficiently cooled for operation, even under atmospheric conditions, the whole internal part of the **motion simulator** is kept at atmospheric pressure with an air venting system. This also results in demanding requirements regarding the tightness of all moving joints and axles in direct contact with the chamber vacuum. The sum of all those requirements makes the construction of a motion simulator one of the most complex engineering challenges of a space simulation facility.

8.3.5 EMC and Magnetics

8.3.5.1 Electromagnetic Compatibility

Aim and Purpose

No space flight system can be managed without electronic and electromechanical components fulfilling diverse complex functions. Besides the intended electrical signals, often **unintended emissions** are sent out which may disturb other components or subsystems in their function. In addition, strong high-frequency **electromagnetic fields** exist which guarantee the communication with the ground station, transfer telemetry signals or measure radar signatures. When

talking about **electromagnetic compatibility** (EMC) the conflict "disturb or be disturbed" is considered. In order to guarantee compatibility, EMC must keep the **unwanted emissions** below a given limit on the one hand, and guarantee immunity to disturbances from the outside (susceptibility to interference) on the other hand. The bigger the margin between the emission and immunity, the bigger the reserves for compatibility. These two values are determined by different methods of testing in an EMC laboratory.

Disturbing Signals and Their Propagation

Components working with clock pulses or switching frequencies (e.g., DC/DC converters, personal computer components) are known to be typical sources of disturbance. Power electronics with fast switching operations (pulse width controlled actuators or motors) are also known as a source of disturbing interference. Fast switching pulses and high currents produce spectra up to high frequencies. Switching transients in the electrical supply system affect all connected subcomponents. **High-frequency disturbances** can travel via the cabling (**conducted emission**) and be radiated either directly from the housing or the cables as an electromagnetic field (**radiated emission**). Natural effects can also generate disturbance signals, such as the **discharge of static electricity** (ESD) or a stroke of **lightning** into the launch vehicle shortly before launch.

When thinking of immunity it is necessary to assess all components and systems as to their susceptibility to becoming interference sinks. Experience shows that, for example, components which convey small analogous signals (signal amplifier, signal conditioning, scientific experiments) are regarded as extremely susceptible to external influences. Caution is advised wherever interfaces lead to the outside and come into the range of transmitting antennas. Conducting susceptibility tests on single leads, and cable bundles, as well as radiating susceptibility tests are undertaken in immunity testing.

EMC Analysis and Procedure

The objective of the analysis is to determine those requirements for the system and the subcomponents which guarantee EMC. These requirements are listed in a specification together with other mechanical or space simulation requirements, normally starting by listing the outer high-frequency fields (**extra EMC**) and taking not only the test object itself but also the equipment of the launcher into account. From these fields a possible "threat scenario" is derived by taking the shielding attenuation of the outer casing into consideration. Furthermore, the **immunity level** is determined. Conversely, the emission limits for the subcomponents are determined, starting with the inner components and the radiation to the outside.

Regarding the compatibility of the systems among themselves (**intra EMC**), basically propagation via lines is responsible. It is necessary to analyze signals in more detail and to find a reasonable compromise with reference to the reduction of emissions and the increase of immunity. This process proceeds mostly in an iterative way by means of experimental results and experience.

EMC Test Methods and Equipment

As stated above, the methods can be subdivided as being for **conducted and radiated emission** and for **conducted and radiated susceptibility**. In physics, each signal can be described by a function in the time domain as well as in the frequency domain. Both methods are used in EMC testing procedures.

Individual **coupling devices (transducers)** belong to all test methods. They establish a connection between the measuring instrument and the physical value to be coupled to or from the unit under test. Many transducers have frequency-dependent factors. Their values have to be calibrated regularly and taken into account when the signals are converted.

The central part of an EMC laboratory is a shielded chamber (see Figure 8.3.20). Considering that our environment is almost everywhere extensively supplied with communication services of all kinds (e.g., radio and television broadcasts, mobile telephones), all of which use high-frequency signal transmission, it must be guaranteed that their signals do not disturb an EMC test. For this reason the object to be tested is placed in a shielded chamber consisting of a metal cabin which does not allow radiation in any direction. It can thus be assured that the signals measured inside the facility actually come from the test object itself and not from the environment. High-quality chambers achieve more than 100 dB screen attenuation up to 18 GHz. A precondition is the filtering of all supply and communication lines.

The test method of **radiated susceptibility** is performed with radiated electromagnetic fields. Without additional provisions the metallic surface

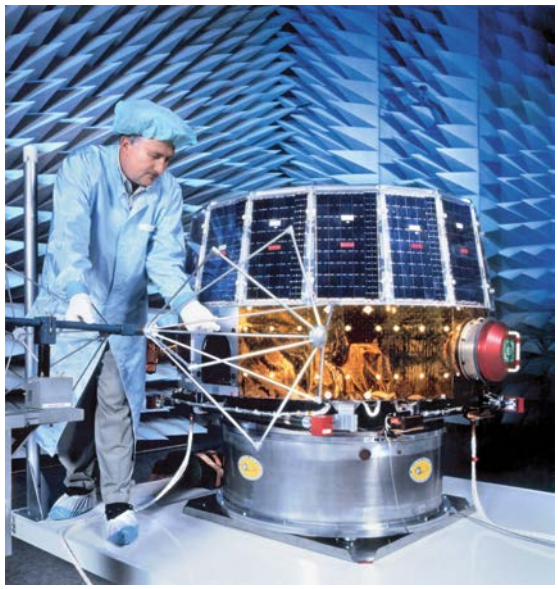


Figure 8.3.20: *Satellite EQUATOR-S in the EMC laboratory (Source: IABG).*

of the shielded chamber would reflect the electromagnetic waves, resulting in standing waves with field maxima and minima. This field distribution does not correspond to the outdoor far-field condition and would falsify a test. In order to suppress wall reflections, the inside of the shielded chamber is covered with absorbers which absorb radiation and convert it into heat. Under electromagnetic considerations these are the given conditions of an infinite space and thus of homogeneous field distribution.

The **thermal efficiency** of absorbers is limited. This feature has to be considered in a risk analysis and is particularly relevant whenever the strong high-frequency sources of the test object are switched on (e.g., the radar antenna or the communications antenna of a satellite).

8

8.3.5.2 Magnetics (Magnetic Cleanliness)

Mission-critical instruments (e.g., magnetometers) on-board scientific satellites in particular need to be protected from spurious magnetic DC fields generated by the satellite itself or its components.

This is typically achieved by running a so-called **magnetic cleanliness program**. This program minimizes the magnetic sources on-board a satellite by employing the following strategy:

1. Wherever possible, avoid the use of ferromagnetic materials.
2. As early as possible, identify the magnetic sources.
3. Characterize all identified sources by modeling and measuring their magnetic behavior.
4. Create a magnetic model of the satellite based on the characterized sources.
5. After assembly of the satellite, measure and calculate the influence of all magnetic sources on the instrumentation.
6. If necessary, try to minimize the magnetic field at the location of the instruments by compensation.

In order to characterize the **magnetic effects** of a satellite or its components, a distinction between hard and soft magnetic effects is necessary. Hard magnetic effects can be directly traced back to the component; the soft magnetic effects on a component are induced by the Earth's magnetic background field. A special test setup allows measurement of the component in a magnetically compensated area and therefore simulates on the ground the magnetic environment of space. IABG operates a **magnetic field simulation assembly** (MFSA) as part of the ESA coordinated test facilities, providing the test environment required to characterize components up to a complete satellite during a magnetic cleanliness program.

Figure 8.3.21 shows the setup of this facility schematically. The use of a special constellation of Helmholz coils allows suppression of the magnetic field of the Earth and therefore the generation of a space free of magnetic fields.

Magnetic Cleanliness Program

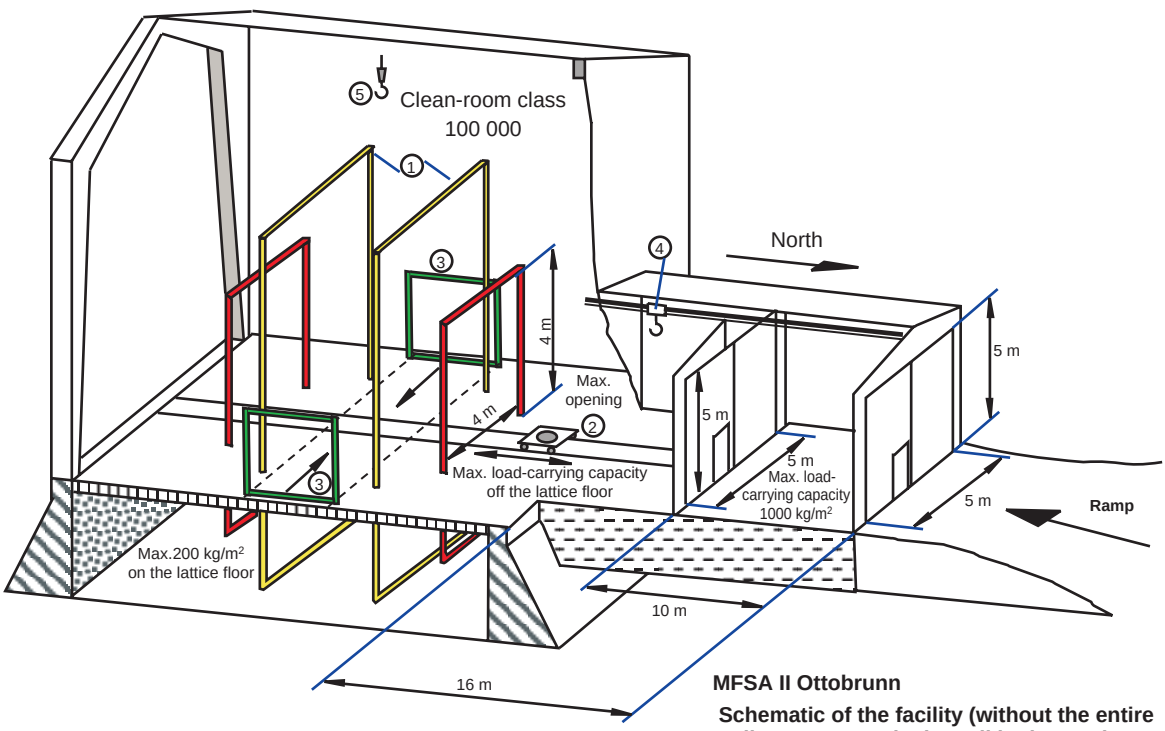
Nearly all magnetic field experiments on-board a satellite suffer from the contamination of DC or low-frequency magnetic fields generated by the spacecraft itself. This problem is known as the **magnetic cleanliness** (MC) problem. It should be noted that the noise of the magnetic field generated by the subsystems of a satellite from DC to some 10 Hz is not covered by the standard EMC program or magnetic verification plans, which typically start at 50 Hz. The EMC program also does not recognize any magnetized elements of the structure and the harness which may cause "DC contamination" of the magnetometer measurements. For these reasons a special program

for evaluating the frequency range from DC up to low frequency has to be employed and the different measurement techniques compared as to EMC.

Magnetic cleanliness programs were successfully run during the Giotto, Cluster, Cassini–Huygens and CHAMP satellite projects. At the beginning of the magnetic cleanliness program a design goal is specified (e.g., 0.25 nT of maximum magnetic distortion at the location of the outermost magnetometer). Based on this design goal a maximum **tolerable magnetic moment** for all individual units is specified. The calculation of the magnetic field around a test object is generally driven by the postulate that a given **static magnetic field configuration** $\boldsymbol{B}_S$ of a test object can be described by a discrete number ($i = 1, \ldots, N$) of single magnetic dipoles with a magnetic moment of $\boldsymbol{M}_i$ at a distance of $\boldsymbol{r}_i$. Written as a vector this yields

$$\boldsymbol{B}_s = \frac{\mu_0}{4\pi} \sum_{i=1}^{N} \left(3 \frac{\boldsymbol{r}_i \left[\boldsymbol{M}_i \, \boldsymbol{r}_i \right]}{\left\| \boldsymbol{r}_i \right\|^5} - \frac{\boldsymbol{M}_i}{\left\| \boldsymbol{r}_i \right\|^3} \right) \tag{8.3.11}$$

Based on high spatial resolution magnetic measurement samples around the test specimen, a **mathematical model** can be constructed which minimizes the differences between the model predictions and the measured field.



MFSA II Ottobrunn

Schematic of the facility (without the entire coil structure; only the coil in the north–south direction is shown)

1. **Center of the coil with a diameter of 6 m**
2. **Trolley with revolving platform**
3. **AEMA (magnetizing and demagnetizing coils)**
4. **Crane: 1500 kg, 5.1 m height of lift**
5. **Stationary crane: 1000 kg in the center of the facility**

Figure 8.3.21: *Magnetic field simulation facility (Source: IABG).*

Therefore, every magnetic control plan contains so-called **rotational measurements** for all magnetically relevant subsystems. The magnetic field of the unit to be characterized is scanned during these measurements. The unit is placed on a turntable and rotated horizontally around the vertical axis of the turntable. A mathematical **multidipole model** is calculated based on these measurements. This model represents the best-fitting approximation between measurements and model. The results of these measurements are combined with the geometrical information about the origin and orientation of every unit on-board the satellite and used to build up a magnetic system model of the satellite. This model gives an idea about which unit contributes most to the magnetic field at the position of the magnetometer on the satellite and the strength of the part of the magnetic field that is compensated by different units of the satellite. Therefore, this model helps to decide whether or not to compensate the magnetic field at the position of the magnetometer.

The **system model** typically gives only a rough estimation of the magnetic fields expected. This behavior is related to the model's limits. It is not always able to predict the mutual interaction of different components correctly and of course it is always possible that not all magnetically significant units are identified and added to the model. Therefore a so-called **magnetic system test** is performed to verify the system model by measurement. From this point on, monitoring of the complete magnetic environment of the satellite starts by placing magnetometers close to the satellite and recording the readings. This monitoring is carried out continuously during integration and until the launch of the satellite. The magnetometers are configured so that they trigger an alarm if defined thresholds are exceeded.

8

8.3.6 Special Environmental and Functional Tests

8.3.6.1 Special Environmental Tests

Depending on the mission, it might be necessary to perform special environmental tests in order to cover special needs. Some examples using current projects are listed below.

Microvibration

Zero gravity offers perfect conditions, for example, for investigating new technologies in materials research or examining the Earth with very precise equipment. Current examples are materials research on the ISS and the exploration of the gravity field of the Earth in the context of ESA's Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) project. These missions demand the absence of **disturbing accelerations** ranging from several milli-*g* down to the micro-*g* region. Over the years it has been discovered that the **reasons for microvibration** are manifold. Besides the active elements on board a spacecraft (mechanisms and altitude control, the crew), mainly passive components such as the satellite bus, solar array or antennas tend to cause microvibration as a consequence of thermal loads, due to the spontaneous release of **thermal stress**. Even the thermally induced compensational movements of the insulation foil (MLI) may cause microvibration to a disturbing extent.

In order to keep microvibration under control and be able to estimate it before a mission, **microvibration budgets** are defined in the design phase and broken down to the equipment level. On this basis, so-called **microvibration measurements** are performed in the course of the qualification activities in order to determine the operational vibration behavior. Currently, there are mainly three different examination methods:

1. Direct measurement of the interface forces (**direct interface force method** (DIFM)): The object is clamped onto a seismic block equipped with force measurement devices at its interfaces. The generated forces created by operation of the device in the low-frequency range (typically 0.01 Hz up to some 10 Hz) can be determined directly at the interface. These interface forces can then be directly used to estimate the disturbing vibrations for the whole system.
2. Indirect measurement of the interface forces (**indirect interface force method** (IIFM)): In the case of higher frequencies, the interfaces can no longer be regarded as being stiff. Therefore, the examined devices are suspended under free–free conditions and

the spectra of the interface accelerations are measured. In order to verify the interface forces, the transmitting functions at the interface patches are experimentally determined and together with the measured acceleration spectra converted into force spectra. The frequency range of this method is from some 10 Hz up to 500 Hz, and sometimes even up into the kilohertz range.

3. **Thermal microvibration tests** gain more and more interest in the context of the choice of suitable MLI foils. With this method, foil specimens up to a square meter in size are attached to a plate with known eigenfrequencies. This fixture is suspended at low frequency in a thermal vacuum chamber and exposed to space conditions. By means of infrared lamps the foils are exposed to thermal conditions. Their behavior is then registered optically as well as with very sensitive accelerometers mounted on the plate. The thermally induced movements of the MLI excite eigenfrequencies in the plate, thus giving an insight into the activity of the MLI foil. The power spectral density of the background vibrations required for these examinations should not exceed $1 \cdot 10^{-12} g^2/\text{Hz}$.

Special Optical Measurements of the Linear Expansion of Camera Structures

For camera structures on Earth observation satellites, essential performance parameters are the expanding coefficients of the structure, which should preferably be very low. A change in the length of the structure caused by temperature differences has an impact on the focal distance and thus on the quality of the resulting images. In order to specify the length expansion of a structure, highly precise optical measurements of the structure are performed during a thermal vacuum test. The measurement accuracy to be achieved is usually within $\leq \pm 2.8\,\mu\text{m}$. In order to achieve such high accuracy it is necessary to eliminate any disturbance in the broader environment of the facility created by the surroundings, such as movement of masses by pumps, air-conditioning, motors, air, persons, etc.

8.3.6.2 Functional Tests

Functional tests are performed in order to verify the proper functioning of the respective system or of several systems together. There are a wide variety of functional tests depending on the mission and they can be adapted to meet special requirements. They can be categorized into three groups, as follows.

Independently Performed Functional Tests

The independently performed functional tests are usually very complex and extensive tests performed separately from the environmental tests. Some examples are listed below:

- **Compact Test Range (CPTR):** A CPTR as operated at Astrium in Ottobrunn offers the possibility to carry out **performance tests** on the **antennas** of a complete satellite, including the integrated payload. The radiation characteristics of antennas, particularly those with large reflectors, are usually specified for the far field. As these characteristics cannot be proven precisely enough with short measuring distances, a homogeneous field is created by means of a CPTR with one or more tilted mirrors at a short distance whose characteristics are identical to those of a far field. This makes it possible to determine the high-frequency characteristics of the antenna at a short distance, such as gain, half beam width, sidelobe suppression and cross-polar suppression. For the foreseen illumination of the defined area, for example for the supply of countries or continents with television signals, the parameters mentioned above are the basis for the antenna design.
- **Firing Test for Solid Propellant Boosters:** In order to verify the **functioning of a booster** as well as the **operating parameters** (thrust, internal pressure, firing speed, temperature behavior during and after firing, nozzle control) firing tests are performed. They are very complex because they require test rigs which can absorb the enormous propulsive forces on the one hand (e.g., the Ariane 5 booster) and on the other hand also allow the realization of the mechanical boundary conditions as they occur during flight. Moreover, such a booster test requires approximately the same effort as a real launch. Apart from the functioning of the

booster, additional parameters such as vibration behavior (which constantly changes during firing due to the mass flow) and the interface forces at the connection points to the central stage are investigated. Test rigs for the European launchers are located in Serdinia, Italy, and in Kourou, French Guyana.

- **Launch Rehearsal:** This examination takes place a few days before the actual launch and includes all functions that are necessary for the launch. In principle, this is the performance of a whole launch sequence without the actual lift-off and serves as a validation of all functions and preparation measures for the launcher, the payload and the ground segment. This procedure allows possible errors to be detected in advance and corrected in time.
- **Engine Test Rig:** At engine test rigs, chemical engines can be operated in order to verify the functioning, the duration of burning and the thrust achieved. The size of the test stand depends on the size of the engine and ranges (from the European point of view) from several newtons up to the large-sized engines of Ariane 5. The most important test stands in Europe are operated by the German Aerospace Center (DLR) in Lampoldshausen, Germany, and by the SAFRANE company in Vernon, France.
- **Separation Tests:** During the launch of a rocket and the beginning of a mission, several separation processes are carried out (stage separation, dropping of the payload fairing, separation of the payload, deployment of the solar arrays, etc.). The state of the art is to initiate them with pyrotechnics in order to guarantee high precision and the appropriate separation forces. It is very important that the separation processes work flawlessly and homogeneously, and that the separated parts do not collide with the remaining system. This procedure is validated during so-called separation tests.
- **Communication Tests between Satellite and Ground Station:** After the satellite has been released into orbit, it is essential that the communication between the satellite and the ground station functions and that the satellite reacts as planned to all commands from the ground station. This communication test between satellite and ground station is very often performed during the test campaign (after the environmental tests and before shipment to the launch site). The satellite is connected to the ground station via its actual communications equipment in order to perform various tests.

Performance Tests during Environmental Simulation Tests

In performance tests, the relevant systems are verified under the respective environmental conditions. Examples are as follows:

- **Performance Tests of the Electronics:** During the thermal vacuum tests, electronic systems as well as whole satellites undergo a performance test in order to verify flawless functioning during all operating conditions. Various temperature ranges are used for this procedure.
- **Deployment Tests of Antennas and Booms:** The deployment mechanisms for the antennas,

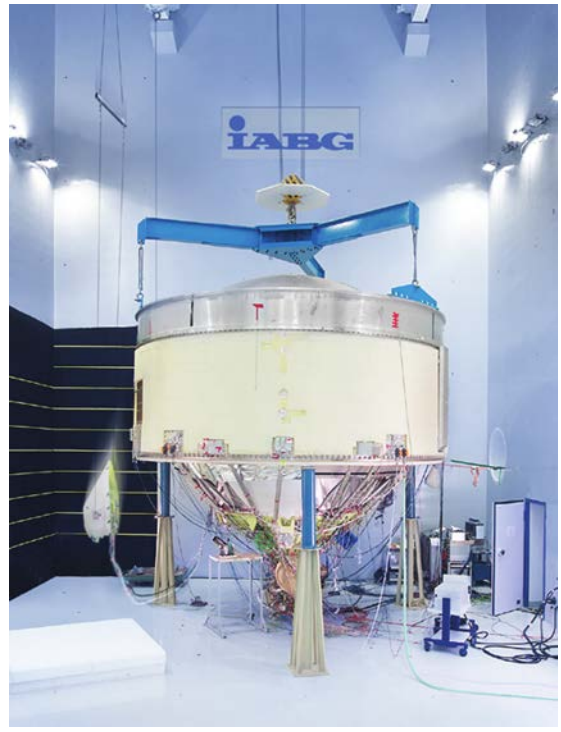


Figure 8.3.22: *Separation test at the Ariane 5 ESC-A upper stage (Source: IABG).*

reflectors and booms are verified under high-vacuum and low-temperature conditions in the context of a thermal vacuum test.
- **Implementing Attitude Control:** Ion thrusters (electrical propulsion systems) as well as gas thrusters are operated during the thermal vacuum tests in order to verify their performance under high-vacuum and low-temperature conditions.

Performance Tests after Environmental Testing
During these tests, systems are verified after having been exposed to the relevant environmental conditions. Examples for this procedure are as follows:
- **Performance Tests of the Electronics:** After the vibration and acoustic test runs, performance tests are carried out in order to make sure that no damage has occurred in the electronics due to the mechanical loads.
- **Deployment Tests of Antennas, Booms and Solar Arrays:** After the vibration and acoustic test runs, the deployment mechanisms are released in order to verify that no damage has occurred to these due to mechanical loads.

8.3.7 Future Developments

In Europe, the construction of **test facilities and test centers** started with the beginning of space flight in the 1960s. Since then, not much has changed in the nature of environmental simulations because the environmental conditions (vibration, acoustic load, electromagnetic radiation, high vacuum, solar radiation and space background temperature) for a satellite in orbit have essentially remained unchanged. The environmental conditions vary only slightly depending on the mission, and only special exploration missions, such as to the Sun or to planets in the close vicinity of the Sun, like Mercury, require specific measures.

Changes in space flight are the result of the performance increase of rockets. The total mass and dimensions of satellites have significantly increased over the years, and test facilities including the infrastructure need to be adapted accordingly. At the same time, the development process of a satellite has shortened significantly. This results in the following facility requirements:
- Increased volumes and satellite masses need to be accommodated.
- Required accelerations still need to be achieved despite increased masses.
- Radiation areas and intensities need to meet requirements.
- Compatibility with the increased number of measurement channels due to the larger dimensions of specimens needs to be assured.
- Operational readiness needs to be guaranteed in a flexible, fast and cost-effective way.

It is expected that this trend will continue in the future. However, it is also expected that the simulation of tests will increase.

Simulation techniques, theoretical models of space vehicles (e.g., via multiple use of satellite buses) as well as computational capacity are continually improving. Any simulation yields errors to some extent, and those errors can only be determined via testing. The same applies to manufacturing defects. Improved theoretical models together with previously simulated tests are important for efficiently conducting tests and understanding the achieved results.

Additional essential requirements for test centers are:
- Test centers must be operated independently from the manufacturer in order to remain unbiased and neutral with respect to the test results and the customer. This independence also guarantees confidentiality of the information gained, such as test results and technical details, drawings, etc.
- Test centers must guarantee that all customers and projects are treated equally, independent of company size and/or project dimensions, and that none are treated in a privileged way.
- Space test centers are required to operate under economic and commercial constraints, respecting decreases in budgets for projects and infrastructure. Another complicating aspect is that a space test center accommodates very maintenance-intensive facilities which usually are single-unit productions.

Bibliography

[8.3.1] ECSS-E-10-03A. *European Cooperation for Space Standardization (ECSS): Space Engineering – Testing.* ESA-ESTEC.

[8.3.2] ECSS-E-10-02A. *European Cooperation for Space Standardization (ECSS): Space Engineering – Verification.* ESA-ESTEC.

[8.3.3] ECSS-E-10-04A. *European Cooperation for Space Standardization (ECSS): Space Engineering – Space Environment.* ESA-ESTEC.

[8.3.4] NASA-STD-7002. *NASA: Payload Test Requirments.* www.standards.nasa.gov.

[8.3.5] Messidoro, P. *et al.* Final Results of the Model and Test Effectiveness Study (MAT€D). 4th SET Symposium, Liege, June 2001, adsabs.harvard.edu.

[8.3.6] Maia, N.M.M. *et al. Theoretical and Experimental Modal Analysis.* Baldock: Research Studies Press, 1997.

8.4 System Design Example: CubeSat

Wilfried Ley, Artur Scholz and Hakan Kayal

8.4.1 Introduction

The mass of a satellite plays an important role because of its direct impact on the launch cost, which is a major cost component of space missions. Especially in the case of interplanetary missions, overall mass reduction is of the highest importance because of the relatively small payload mass compared to other missions. One possible way to reduce cost is miniaturization. The cost for applications which can be realized by satellite missions can be reduced by using lightweight and small satellites. One of the latest activities in the direction of miniaturization is the development of picosatellites. Based on the work of two universities in the USA, namely Stanford University and California Polytechnic State University, a standard for picosatellites, called CubeSat, has been developed. CubeSats are **picosatellites** with a maximum mass of 1 kg and dimensions of $10 \times 10 \times 10\,\text{cm}^3$. The standard is based on a deployment mechanism, the so-called P-POD, which can handle two or three CubeSats, allowing the payload mass to be increased accordingly. Today, more than 60 universities and diverse companies are working on CubeSats. Figure 8.4.1 shows an engineering model of a CubeSat. The applications for this class of satellites are still limited because components which exist for larger satellites are not available for CubeSats. The development of CubeSats is at an early stage.

The attitude control plays a special role here. A well-defined orientation is essential for many applications such as Earth observation or astronomy, otherwise the target of the observation would be missed.

In order to be able to control the attitude it must first be determined which sensors are to be used. For most of the attitude sensors known from larger satellites, such as star trackers, Sun sensors, gyros or Earth sensors, no appropriate versions are available for picosatellites. Existing sensors are too large, too heavy and require too much power. There is a similar situation for attitude control actuators. That is why in most CubeSats today permanent magnets or magnetic coils are used for attitude control. In many cases there is no attitude control at all.

Another bottleneck in the design of picosatellites is communications. Many applications, such as Earth observation, require the collection and transmission of large amounts of data to the ground station. The typical communication bands used today are the UHF and VHF bands, allowing limited data rates in the range of 1.2 to 9.6 kbit/s. Thus it takes a very long time to transmit larger images, which would make such applications practically impossible.

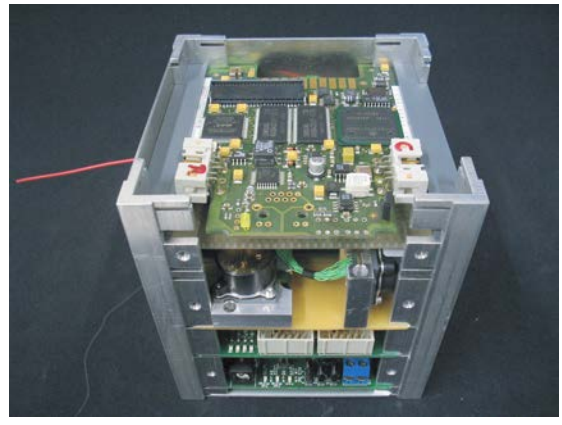


Figure 8.4.1: *Engineering model of a CubeSat (Source: TU-Berlin).*

In order to reduce the lack of such components and to allow more demanding applications, projects are already in progress at different places or new ones are being planned. Two examples are microwheels suitable for picosatellites and an extremely low-energy S-band transmitter.

One of the advantages of **miniaturization** is the potential use of these components in larger satellites, such as **nanosatellites**. S-band transmitters for example can also be used in nanosatellites and would help to maximize the payload capacity remarkably. The same is of course true for other components, such as sensors for attitude determination. This additional payload capacity could then, for example, be used for larger optics, which would enable nanosatellites to run applications for which normally small satellites would be needed.

The number of potential applications for picosatellites will increase with the use of more and more suitable components. Especially promising are applications in conjunction with **constellations** or formations due to the low launch costs. Using constellations, it might be possible to build sensor networks in space where each of the satellites would take measurements simultaneously at different positions in orbit. Some of the application fields can be Earth observation, observation of astronomical objects or space weather. This approach makes it possible to organize a much denser network of sensors than would be possible with larger satellites for cost reasons. It is conceivable to use either the same or different types of sensors on each satellite in the constellation, which also increases system redundancy. Moreover, it might be possible in the future to combine the sensors of satellites within a formation to increase the aperture of an observation system.

Another field of application is to use picosatellites for **exploration of the Solar System**. Picosatellites could, for example, be ejected from larger spacecraft orbiting a planet and support the tasks of a probe, a surface-bound system or a rover.

An additional application field for picosatellites is the verification of miniaturized components which, as mentioned above, are becoming more and more important. Miniaturized components can be **verified in orbit** much more frequently and at much lower cost by using picosatellites.

8.4.2 Mission Concepts and Scenarios

A typical CubeSat mission, like any other satellite mission, consists of the following mission elements:

- Mission objective
- Payload
- Satellite bus
- Orbit
- Launch segment
- Ground segment.

There are many similarities with respect to the systematics and tasks of the mission elements. However, the extent of these elements must be adjusted to the constraints, which in the case of CubeSats mostly mean that there is a very limited budget, limited time and a lack of experience on the part of the developers. Thus the first steps in establishing a CubeSat mission are to define the time frame and budget and to define the mission objectives.

8.4.2.1 Mission Objective

The mission objective is the real task, the actual meaning of each CubeSat mission. The expectation for many scientific or technological applications is to verify the satellite or newly developed components for reliable operation in space, which opens the door for use in commercial applications. This so-called **technology demonstration** is the declared objective of most CubeSat missions.

Not all CubeSats are limited to technology demonstrations. Many missions incorporate a camera to make images of the Earth. There are also efforts to equip CubeSats with instruments to measure remnants of the atmosphere in space, or to measure changes in the Earth's magnetic field for clues on how to forecast earthquakes. But of course other objectives, such as communication tests with high data rates, robotic missions, constellation flights and Moon missions, can also be imagined.

8.4.2.2 Payload

The payload is directly correlated with the mission objective, because it is only possible to reach the mission objective by using the payload successfully. In most cases it is an instrument or a sensor which generates

essential data for the satellite mission. The very strict constraints regarding mass, volume and power drastically limit the choice of available instruments. That is why CubeSats have been rarely considered so far as an alternative by scientific organizations. Thus, there is a large demand for the **miniaturization** of instruments.

If the mission is a pure technology demonstration mission, then the component to be tested is the payload, which could be a component of a subsystem, a subsystem or even the satellite itself. In the last case, there is no simple distinction between satellite bus and payload.

8.4.2.3 Satellite Bus

To achieve the mission objective a payload is needed, and to operate the payload safely in orbit a satellite bus is required. The tasks of the satellite bus are the mechanical integration of the payload, data communication with the ground station, thermal and attitude control of the satellite, and other tasks described in the system design section.

8.4.2.4 Orbit

Up to now only **low Earth orbits** (LEO, 300–800 km) have been considered suitable for operating CubeSats. There are several reasons: one is that the power demand of the communication subsystem for reliable data transmission increases dramatically with increasing orbital height, something which cannot be handled by today's CubeSats. Moreover, higher orbits are more demanding as to radiation load, thermal issues and power management. Finally, the launch cost for LEO is lower than for higher orbits.

The mission objective and mission operations determine the selection of a suitable orbit. A Sun-synchronous orbit, for example, can be very suitable regarding the power and thermal subsystem requirements and provides a nearly global coverage. On the other hand, an equatorial orbit is more suitable for attitude stabilization with a permanent magnet since a more homogeneous magnetic field exists there.

8.4.2.5 Launch vehicle

In order to reach orbit a suitable launcher is needed. In principle a CubeSat can be launched by any modern launcher as the **vibration test specifications** for CubeSats (and thus the qualification for flight) are satisfied by virtually all present launchers. However, currently there are not many launches of CubeSats. One reason is that the operators of the main payload (mostly commercial satellites) must accept the CubeSat as a **secondary payload**. The launch service providers too must accept a secondary payload, and often it is not worth taking the risk. Luckily, two to three CubeSat launches a year have taken place, thanks to the efforts of institutes in Canada, the USA and Japan. The former Russian intercontinental ballistic missiles (ICBMs) like DNEPR and Rockot are especially suitable for these launches, and there are also efforts to develop new launchers for small satellites. But in that case the launch costs would again be very high. There is another, simpler option for launching CubeSats, which is direct ejection from the Space Shuttle.

8.4.2.6 Ground Segment

The ground segment incorporates the terrestrial infrastructure and the personnel necessary to operate the satellite safely. Building up and operating an amateur radio ground station is affordable for private persons and for universities. Students can operate such a ground station as well as perform the data analysis.

8.4.3 Requirements

The main reason for the popularity of CubeSats is especially the standardized interface to the launcher, namely the **P-POD** (Poly Picosatellite Orbital Deployer). Because of the specifications defined in this standard, the satellite is widely decoupled from integration in the launcher. This makes it possible to start the development of the satellite before the launcher itself is selected and financed. But the CubeSat developer then has to fulfill some strict requirements. These include the following for each CubeSat (see Figure 8.4.2):

- The structure is a cube with dimensions of $100 \times 100 \times 100\,\text{mm}^3$.
- Anodized edges with a length of 113.5 mm in contact to the P-POD.
- Overall mass less than 1 kg.

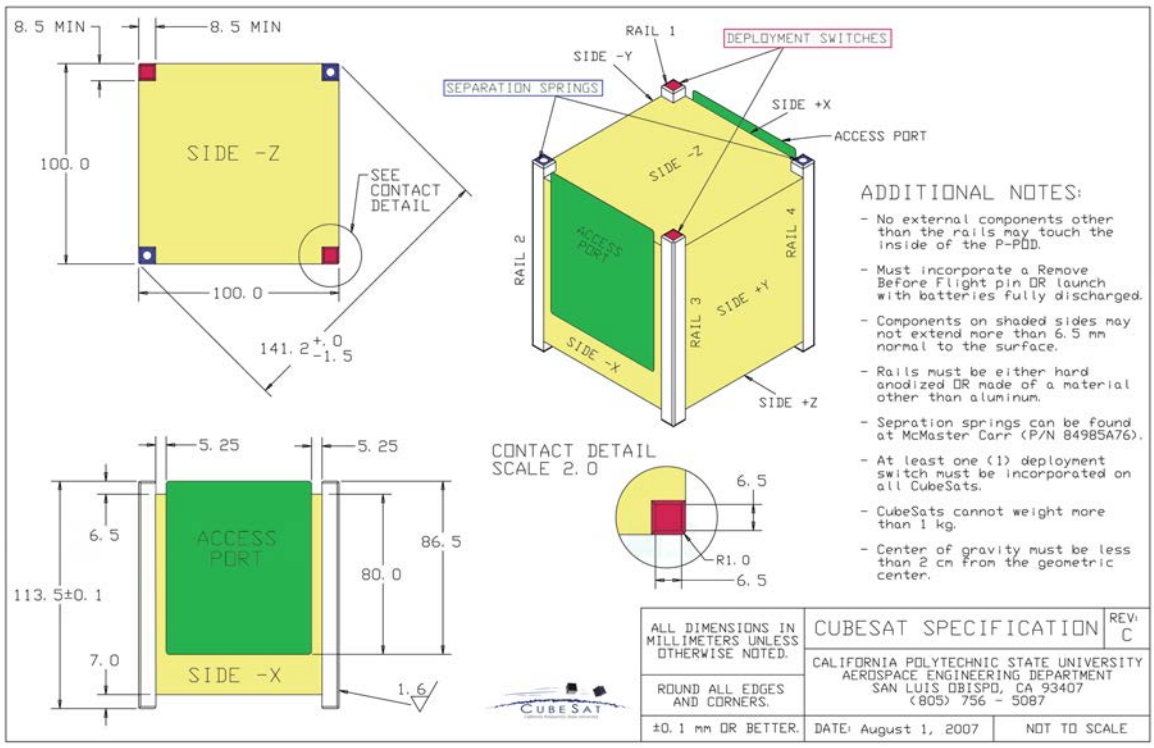


Figure 8.4.2: *CubeSat specification (Source: California Polytechnic State University).*

- No pyrotechnical elements.
- Integration of a transmitter shutdown command in order to switch off the transmitter from the ground.
- Other requirements in order to achieve smooth interaction between the P-POD and other CubeSats contained in it.

The standard has been extended so that double- or triple-sized CubeSats are possible combining the volume of two or three units.

Additionally, it is required that the CubeSat pass **vibration and thermal tests** in order to ensure that it does not endanger other CubeSats within the P-POD and the main satellite on-board the launch vehicle. The important aspect here is to minimize any risk for the main payload which may prohibit the CubeSat(s) from sharing the flight.

Once this is ensured, it is left to the developers of the CubeSats to decide how far they want to go to ensure the functionality of their satellite in orbit by testing it on the ground. Because the missions so far are noncommercial and have a pure research background, requirements for redundancies and extensive tests integration are often waived in order to keep the development and test costs low.

8.4.4 System Design and Subsystems

Apart from the standard specifications which are binding for CubeSats, each development team is free as to the design of the satellite. In other words, CubeSats may look similar from the outside, but inside each satellite the individual expertise and different approaches of the developers can be found.

But there is one big commonality among all CubeSats, and that is the use of almost exclusively commercial off-the-shelf (COTS) electronic components. This is done in order to keep the costs low, because

space-qualified parts are extremely expensive. One positive side effect is that these parts often have better performance related to mass and power demands compared to their space-qualified counterparts. This is because the latter have to be qualified by design or tests, which takes a lot of time. In fact, such chips are already out of date when they are released on the market. However, the space branch is very conservative and mostly only such parts are used. Thus CubeSats can turn this situation to their advantage by using the latest hardware at the cost of higher risk.

For the **design** of a satellite it is useful to build the satellite bus around the payload. This means that the payload and the mission objectives are the design drivers for the subsystems. The subsystems are generally divided into the categories described below. The requirements for each of the subsystems can be derived directly from the above two criteria.

The following sections give a short overview of the methodology and technology for realizing each subsystem of a CubeSat. Underlying concepts are introduced and future developments are outlined, although the discussion is not complete due to the limited framework of this handbook.

8.4.4.1 Attitude Control System

Currently, either no attitude control or only very basic attitude control systems are used in almost all CubeSats. The main reason is that there are basically no suitable actuators and sensors available for CubeSats. Sensors and actuators which are known from larger satellites are too large, too heavy and need too much power to use in CubeSats.

The simplest form of controlling a CubeSat's attitude is to use permanent magnets. This stabilizes the attitude of the satellite so that fast rotations can be avoided. But it is not possible to achieve a controlled attitude change with this system. Attitude control systems using magnetic coils as actuators are an enhancement. Attitude determination is then often accomplished by magnetometers or by using the solar array currents. The expected accuracy of such a system is very low.

There are, however, efforts which aim to develop components such as picoreaction wheels or miniaturized sensors. Figure 8.4.3 shows such a system of

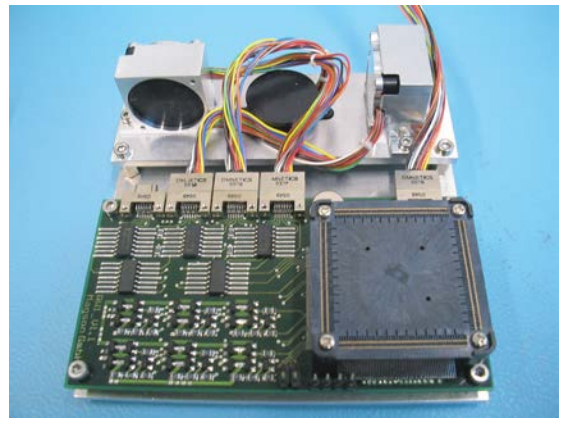


Figure 8.4.3: *Prototype of a microreaction wheel system for CubeSats (Source: TU Berlin).*

reaction wheels together with its electronic module. The aim is to achieve an accuracy similar to what is possible today with nanosatellites. Once this is achieved, it will be possible to have applications for picosatellites which require precise attitude control.

The complexity of the attitude control system of a CubeSat (e.g., three-axis stabilization) does not differ much from the larger satellites. The requirements regarding functionality are the same. Only the requirements for redundancy and reliability can be reduced in order to remain within the cost frame. On the other hand, more than one CubeSat might be built, which could provide redundancy on the system level.

8.4.4.2 Propulsion System

There are also no specific micropropulsion systems suitable for CubeSats. But two different methods are being developed: **chemical and electrical propulsion**. Cold gas propulsion systems or, for example, systems using hot water vapor belong to the first category. These systems are able to produce significant thrust for a short time in comparison to electrical propulsion systems and are therefore more suitable for CubeSats. Electrical propulsion systems use electrons or ions for impulse generation, which provide low but continuous thrust levels. With such a propulsion system a reentry system could be designed which ensures the reentry of the satellite after a defined time period. Another possibility would be to increase the orbital altitude

by initiating a spiral path which would transport the satellite upward until it reaches a lunar orbit.

8.4.4.3 Communication System

In order to communicate, the satellite must be able to receive telecommands and send data. Because computer systems communicate using **binary data**, whereas a radio communication link uses **electromagnetic waves**, a number of components must be incorporated into the communication system. In the following, data transmission from a CubeSat to a ground station is explained as an example. This discussion is principally limited to the use of amateur radio frequencies, due to the high complexity and costs associated with registering a dedicated frequency.

The data to be transmitted (e.g., measurements, status information, etc.) is first formatted on-board the satellite by the board computer, and packetized if there is a large amount of data to be transmitted. The **AX.25 Protocol** is very suitable for this purpose and widely used amongst amateur radio operators. But in principle it is possible to use one's own protocol. Then the binary data is modulated by a modem into a baseband signal of lower frequency. Often the slow AFSK 1200 bit/s modulation scheme is used, which can be demodulated by many ground stations. For a better bit error rate it is advisable to use MSK, PSK or some other modulation technique with higher bit rates. The baseband signal is then modulated onto a high-frequency carrier signal by the transmitter. This is achieved by using FM modulation by commercial transmitters, which are used in CubeSats. Finally, the signal is transmitted into space as an electromagnetic wave by the antenna. In general some kind of rod or patch antenna can be used for this purpose.

Uploading of commands is done in a very similar way. The difference is that on Earth the ground station may use much higher power than the satellite does for transmission. On the other hand, the amount of data to be transmitted to the satellite (telecommands) is a fraction of the data which the satellite has to transmit. In this case it is more important to have a reliable link than a high bit rate. One possible solution is to use DTMF tones for commanding (of course again modulated onto a high-frequency carrier signal).

Almost all CubeSats transmit a **beacon signal** which contains the call sign of the satellite and critical status information. This is often done in Morse code utilizing OOK (On–Off Keying), which switches the carrier on and off. The beacon is also used to better track the satellite.

8.4.4.4 Power System

The only energy source which can so far be used for CubeSats is the conversion of sunlight by solar cells. As the total area on which solar cells can be mounted is very small due to the small size of a CubeSat, the conversion must be done very efficiently. New developments (**multijunction GaAs solar cells**) are preferred for this reason, although their price is much higher and they are more difficult to purchase than other options. If the satellite is planned to be active in the Earth's shadow, which is desirable, for example, when recording system status information, then batteries must also be built into the system. In this case modern **lithium-ion accumulators** are used because of their high energy density. An integral part of the design of the energy system must be to ensure that the satellite has enough excess energy in the sunlight phase (which is used to charge the batteries) to compensate the loss in the shadow phase after considering all losses in the system.

Another element in the power system is the conditioning and distribution of the energy. Using DC/DC converters (upconverters or downconverters) the voltage levels are transformed from the solar cells or batteries to the levels needed for the power bus systems. The most common voltage levels are 3.3 V and 5 V as the power supply for electronic circuits. Selected loads or complete subsystems can be switched on and off by using transistors. This can be done, for example, to reset the board computer or to put the satellite in a power-saving mode.

8.4.4.5 Command and Data Handling System

The on-board computer is the central unit for processing digital procedures and data on the satellite. In some cases CubeSats can incorporate more than one computer system to fulfill different specific tasks. But all high-level information is also routed through

the main computer, a microcontroller. The selection of a suitable MCU (Micro-Control Unit) is arbitrary as there are many modern MCUs and a high number of derivatives provided by established manufacturers. The 8051 and HCS12 families of microcontrollers are widely used. It is of course important that the MCU has all the needed interfaces and enough I/O pins. One of the interfaces, for example, is the I2C bus interface. This bus system is widely used in entertainment electronics but it is also suitable for CubeSats due to its simplicity. With this bus system the controllers can communicate and also read out other components and control them. Besides receiving and executing commands from the ground station, the board computer has to read and store the necessary data. Temporary data can be kept in a SRAM, while data which have to be stored for longer time periods or large amounts of data are stored in a flash-type memory. This type of memory retains the data even if the power supply is switched off. However, the radiation issues regarding flash memory are controversial, so it is a good idea to plan suitable means to protect the data (e.g., by using error-correction codes).

8.4.4.6 Thermal System

In a typical LEO with about 60 minutes of sunlight and 30 minutes of shadow a CubeSat is exposed to extreme temperature changes. In the shadow the satellite is exposed to almost space background (approx. −270 °C), while in the sunlight it is exposed to one solar constant plus the reflected light from the Earth's atmosphere or from the Earth itself (albedo). As a rule of thumb (confirmed by real measurements of CubeSats) it is assumed that the temperatures inside the CubeSat will be between −20 and +20 °C. But when the area is directly illuminated by the Sun the temperature is expected to be +80 °C. Although these temperatures are relatively moderate (most of the industrial parts survive these temperatures without any problem), for some parts special measures must be taken. This might be a measurement device or the battery, which cannot operate at low temperatures. It is often not enough to use passive elements (paint, surface structure, etc.) to prevent these elements from undercooling. In these cases it is mandatory to heat them actively. Heating foils are suitable for applying heat directly. But because such heaters consume a lot of energy it is advisable to carefully plan the optimum configuration and location so that the duration of heater operation can be minimized.

8.4.4.7 Structure and Mechanisms

As mentioned in the introduction, the outer structure is defined by the CubeSat specification. In addition, the specification requires the design of a so-called "access port," which is used as access to the satellite to "remove the before-flight pin," which enables the power system of the satellite. The real challenge in the design of a CubeSat lies in the design of its mechanisms. The tight conditions of the outer dimensions do not allow large parts on the outside. External parts are allowed to extend only 6 mm at maximum in order not to touch the P-POD walls. Mechanisms such as antenna or solar panel deployment must live with this limitation.

Because of the small size, the inner structure can also be very compressed. It is not enough to shrink the components and then build the satellite in a box-like architecture. It may be necessary to distribute functions which are normally built into one box over the different levels of the satellite. Integration and cabling can therefore be demanding challenges from the construction point of view. The CubeSat specification allows enlargement of a single CubeSat to a double or triple unit. This possibility is used to get more room for the components and to increase the power capacity.

8.4.5 Model Philosophy

The development of the final flight model of a CubeSat goes through a series of test models, which become more and more detailed from one project phase to the next. These are functional prototypes, engineering models, qualification models and finally the flight model. Because of time and budget restrictions, the qualification model and the engineering models are often combined, as described in the following. Combining an engineering model and a flight model (building, testing and flying the engineering model) is, however, not advisable since the engineering model

may comprise many items which have to be revised because of lack of experience.

8.4.5.1 Prototypes

Prototypes are the first steps to verify the function of the satellite. The aim here is simply to verify the function of the design of new parts or assemblies.

If, for example, a battery which has no space flight heritage and is not qualified for space is used, it must first be tested extensively. In this case the battery would be exposed to vacuum for long time durations and many charge and discharge cycles would be applied to it. Then, when the satellite is constructed and all systems integrated, the satellite undergoes extreme environmental tests in the assembly, integration and test (AIT) phase. For this reason the parts or assemblies must be tested in advance to ensure that they will function later when they are installed in the satellite. It is extremely important of course to verify not only that the components survive the environmental conditions, but also that each system fulfills the task it was designed for.

For this purpose so-called **breadboards** are built for electrical assemblies. These are supplied with all necessary inputs (power, data, measurements, etc.) according to the black box principle and the outputs are monitored. Even circuits developed by IC manufacturers are recommended to be built up and tested using this method. The prototypes of the different assemblies (Figure 8.4.4) can be successively integrated afterwards to test the functionality on the subsystem level. The test on the system level then takes place during the AIT phase.

8.4.5.2 Engineering Model

After successfully concluding the tests with the prototype, the results of which are discussed at the **critical design review** (CDR), the development continues with the construction of the **engineering model** (EM). All design features and interfaces are frozen at this point, which means that no further design change is allowed, unless it is absolutely necessary. As mentioned in the introduction, the engineering model should be designed to resemble as closely as possible the functions of the future flight model (FM). For this purpose the EM must have the mechanical and electrical properties of the later flight model.

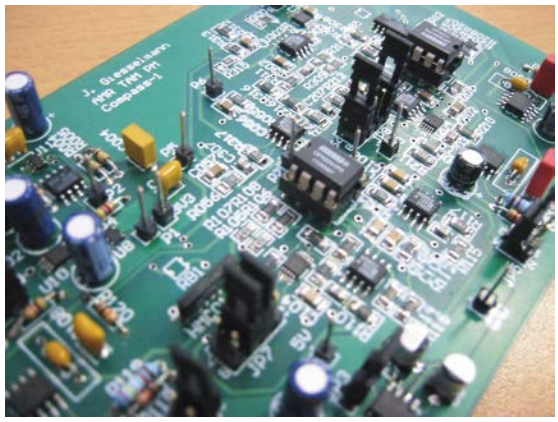


Figure 8.4.4: *Magnetometer at the prototype level (COMPASS-1) (Source: FH Aachen).*

8.4.5.3 Flight Model

In the ideal case the engineering model would be suitable as a flight model. But normally small differences arise, especially during the qualification tests in order to correct minor issues. The test requirements are treated in the AIT section. At this point it should be mentioned that after finishing the EM tests, the flight model can be constructed. Before that there is a **qualification review** (QR) in which the eventually necessary changes for the FM are analyzed and defined.

The model philosophy of the BeeSat picosatellites, for example, includes the following models.

The **structure model** (SM) is used not only for the tests mentioned in Table 8.4.1, but also for the verification of the cabling and integration concept. The **radio link test model** (RTM) is used to measure the antenna characteristics in conjunction with the actual structural properties of the satellite. This is important due to the fact that the satellite is very small and strong interactions between the structure and the antenna characteristics can be expected. The combined **engineering and qualification model** (EQM) goes through all qualification tests. The results can be used to correct any faults for the flight model. If necessary, the EQM is further used at the laboratory for diagnostics after the launch of the satellite into orbit.

Table 8.4.1: *BeeSat verification on the system level.*

	SM	RTM	EQM	FM
Electrical function tests, performance tests		×	×	×
Payload system tests			×	×
Physical properties	×		×	×
Resonance search	×		×	×
Random vibration tests	×		×	×
Sine vibration test	×		×	×
Thermal cycle test			×	×
Thermal vacuum test			×	×
Thermal balance test			×	×
Antenna deployment	×	×	×	×
Antenna gain		×		
Compatibility test		×	×	
Eject test	×		×	×
Mission operations test	×	×	×	×
Software upload tests			×	×
End-to-end system tests			×	×

8.4.6 Assembly, Integration and Test

The AIT phase can be considered as the most critical phase in the development of the satellite. In this phase it must be shown not only that each assembly and subsystem does work properly, but that all parts are working together. All previously defined interfaces must be tested extensively in order to detect and correct possible failures (e.g., endless loops in software). The mission should be simulated with all relevant aspects such as sending telecommands and receiving data, because, although everything may be modeled very carefully, some errors can only be detected in the fully integrated system.

8.4.6.1 Integration

When the development and construction of all assemblies and subsystems is completed, they must be integrated. It is very helpful to write a procedure in advance which shows the integration steps one by one. A CAD model can be used for this procedure but real conditions, for example the accessibility of screws or the sequence of integration, should be kept in mind as well. The integration itself should take place under **clean-room conditions** as far as possible. Compared to larger satellites, where several weeks must be planned for the integration process, it takes only a few hours for CubeSats, provided that it is well organized in advance. It is also very useful to extensively document the integration process with photos. After integration the satellite can be accessed via the access port (or via the radio communication link) in order to receive measurement data or to check specific system functions. When it has been ensured that all systems are working properly, the environmental tests can start.

The aim of these tests, which are described in the following, is the qualification of the satellite for the launcher (whereby it should be ensured that the satellite does not endanger other payloads on-board the launcher and can safely reach the orbit) and to ensure that the satellite will survive the extreme orbital conditions.

8.4.6.2 Vibration Tests

The first group of tests is vibration tests, where the mechanical loads which are induced on the satellite during launch are simulated. Due to the fact that in many cases there is no signed launch contract during the development phase of the CubeSat (in many cases the launch negotiations are started after the qualification review), these tests should be organized in such a way that a variety of launchers can be considered. As an example, the test profile for the NSL-4 launch used by the University of Toronto is given in Figure 8.4.5. To make the simulation as realistic as possible the CubeSat is built into a test-POD (similar to the P-POD), which is then mounted on the vibration table. For all three axes the random profiles are executed, whereby the levels for the EQM are increased by a factor of 1.5 compared to the FM. Sine tests are done before and after each test, which can give additional information on structural changes of the satellite.

8.4.6.3 Thermal Vacuum Tests

With the thermal vacuum tests the conditions in orbit are simulated as realistically as possible. These tests can be very expensive as a number of **environmental conditions** must be simulated correctly. This

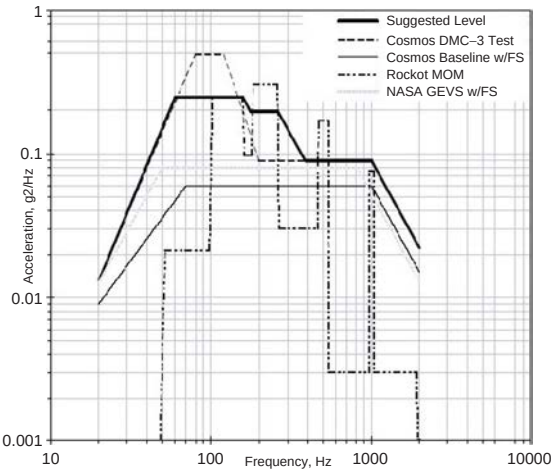


Figure 8.4.5: *Vibration profile (Source: University of Toronto, Institute for Aerospace Studies).*

begins with the **solar illumination**, simulated by a xenon lamp. It must be adjusted to a level where the CubeSat is irradiated with one solar constant. The cold space background is simulated by cooled black walls. In addition, **vacuum** must be generated of course. The small size of a CubeSat is advantageous in this case. A vacuum chamber and a xenon light source lamp suitable for this size are affordable. The **cold background** requires more effort. It is, however, possible to design the thermal analysis in such a way that it can be decoupled from the test. If the "space background" is at room temperature level, then it can be incorporated in the thermal analysis. Accurate temperature predictions can be made afterwards by calculating the conditions in orbit. Another possible way to test the temperature behavior of the satellite is to use thermal ovens. Such thermal ovens are able to provide temperature cycles from degrees below freezing to high temperatures under atmospheric conditions and can be used to expose the CubeSat to extreme temperature cycles.

8.4.7 Operations Aspects and Ground Segment

The operation of a CubeSat does not principally differ from the operation of a larger satellite. All known aspects of larger satellite operations must be considered for CubeSats as well. All elements of mission planning and mission operations must be involved.



Figure 8.4.6: *Mission control center of TU Berlin (Source: TU Berlin).*

One important aspect of mission planning is the visibility prediction of the satellite from the ground station. In the case of CubeSat projects it is common today to use suitable freeware or shareware software tools. These programs need so-called **two-line elements** (**TLE**) as input to calculate the orbit of the satellite. TLEs are orbital parameters of satellites and other objects in Earth orbit. These elements are provided by the US NORAD in a two-line test format. The accuracy of the elements is good enough for ground station antenna tracking. Having the access times to ground stations or other points on Earth makes it possible to determine the switch-on times of cameras or transmitters, for example. The planning itself is mostly done by simple means such as using tables.

Mission operations can be performed by one or more ground stations connected to a mission control center (see Figure 8.4.6). In the simplest case the ground station can be used for all tasks, including planning and mission operations.

For the ground station, equipment which has been proven to be useful for an amateur radio via a satellite is employed in most cases. A typical set of equipment includes:

- Yagi antenna
- Amplifier for transmission and reception
- Lightning protection
- Antenna rotor

- Antenna control electronics
- Computer for antenna steering, including orbital tracking software
- Transceiver for UHF/VHF bands
- Terminal node controller (TNC)
- Computer for data reception and transmission of telecommands.

In most cases the AX.25 Protocol is used for the data transmission. The big advantage of this protocol is that it is widely used by amateur radio stations, so that there are a lot of hardware and software tools available which can be used. The connection between transceiver and computer is realized by the TNC. It includes a modem, which is connected to a microcontroller which takes care of the time-critical functions of the communication. The cost of such a ground station is in the order of several thousand euros, which makes it affordable especially for universities.

Bibliography

[8.4.1] CubeSat Design Specification, Revision B. California Polytechnic State University, Aerospace Engineering Department, 2004.

[8.4.2] Wertz, J.R., Larson, W.J. *Space Mission Analysis and Design*, Third Edition, Space Technology Library. Heidelberg: Springer Verlag and Microcosm Press, 1999.

[8.4.3] Sandau, R. (ed.) *International Study on Cost-Effective Earth Observation Missions*. Leiden: Balkema, 2006.

8.5 Exemplary System Design of a Microsatellite Mission

Wolfgang Bärwald and Klaus Brieß

8.5.1 Microsatellite Design Philosophy

Satellites can be classified by means of their total mass. Above all, they can be divided into conventional satellites and small satellites. **Small satellites** can be divided with reference to their mass into micro-, nano- and picosatellites. There is no unified international definition, but the classification in Table 8.5.1 is common.

The first satellite, Sputnik 1, was a small satellite in this sense. But the next satellites became much bigger and the launcher capacity much larger. Today many highly complex satellites are in space. Advanced technologies allow these missions to be supplemented by small satellite missions. New technologies are verified in space by small satellite missions. With the experiences gained, they can be further developed for large satellites. The advantages of small satellite missions are:

- Low budget
- Short mission implementation time
- Dedicated mission objectives
- Small user communities
- Possible implementation of advanced technologies without space flight qualification.

Small satellite missions have to fulfill high performance requirements under difficult conditions: low mass, low volume, low power consumption, low financial budget. The cost aspect is the most important driver for small satellite missions. Usually the main part of the cost of such missions is for the spacecraft itself. The launch costs are about 20–50% of the total mission cost and the costs for the ground segment and mission operations are in the order of 15% [8.5.1].

In the satellite classification no sharp limits exist relating to the mass. Even a 120 kg satellite can be called a microsatellite. A very important feature of the class of **microsatellites** is the **piggyback launch strategy**. This means that the microsatellite can be launched as a secondary payload. That is one of the main differences to minisatellites, which require a more expensive shared launch or a dedicated

Table 8.5.1: *Satellite classification by total mass.*

Satellite class	Mass
Conventional satellite	> 500 kg
Minisatellite	100–500 kg
Microsatellite	10–100 kg
Nanosatellite	1–10 kg
Picosatellite	0.1–1 kg

launcher. But also the launch strategy is not fixed. For instance, the nanosatellites of TU Berlin (TUBSAT-N) are launched into orbit with a dedicated submarine rocket, and a few microsatellites have also been launched with a dedicated launcher developed in-house.

The application of the design-to-cost philosophy is typical of microsatellite missions since costs are such an essential factor. Cost overruns in microsatellite missions lead to the danger of termination of the complete mission even before it begins. This is also a very special feature of microsatellite missions in contrast to larger missions.

In the following the typical system design of a small satellite mission is described using the BIRD microsatellite mission of the German Aerospace Center (DLR) as an example. BIRD was launched successfully into low Earth orbit in 2001.

8.5.2 Design and Mission Elements of BIRD

8.5.2.1 Motivation and Mission Objectives

Fire has an increasing impact on the Earth's ecosystem. It influences the chemistry of the atmosphere, the greenhouse effect (because of carbon-containing aerosol emissions) and contributes to climate change. The need for precise data on the intensity and surface extent of vegetation fires and volcanic activities is increasing worldwide. Such data is important for the detection and management of disasters and for a better understanding of the impact of fires and volcanic activities on the climate. Their influence has not been investigated sufficiently up to now because of the lack of measurement data on a global scale.

For hotspot events such as forest and vegetation fires, volcanic activity or burning oil wells and coal seams, dedicated space instrumentation does not exist. A new generation of cooled infrared sensor arrays will play an important role for future Earth observation missions. For the measurement of temperature, area extent and intensity of fires and volcanic activities from space at least two infrared imaging sensors interfering with each other are needed, preferably in the wavelength range of 4–10 µm. The infrared sensor system should not be saturated at fire temperatures up to 1000 °C. The temperature of the "cool" environment of the fire areas should be measured with a resolution of approx. 0.5 K.

The **BIRD (Bispectral Infrared Detection)** mission is a small satellite mission with technological and scientific–methodological objectives, answering a number of questions related to the evaluation and operation of cooled infrared sensors in space. Another primary mission objective is the investigation of vegetation fires from space by means of this new infrared technology. Innovative operational concepts and new methods of data acquisition for hotspots are demonstrated by this mission. The objectives are summarized in Section 8.1.4.

8.5.2.2 Mission Concept

The BIRD mission is funded and implemented by DLR. Small and medium-sized enterprises and national research organizations like the Fraunhofer Gesellschaft are contributing their own funding to the success of the mission. Strict cost limits, very challenging mission objectives and the implementation of advanced technologies also led to new approaches in the mission program segment. The use of advanced technologies applying the design-to-cost philosophy was feasible only by keeping the cost of parts and equipment low, which meant the general abandonment of so-called "high-reliability" space-qualified parts. To reduce the resulting higher risk, the following measures were taken:

- Designing for overloads and smart degradation, when operating with underperforming satellite subsystems (derating parts and equipment).
- Extensive redundancy conception.
- A tailored hybrid model philosophy on the system level (2½ model philosophy: STM → EM, FM).
- A tailored quality management system according to ISO 9000 and successive standards, ECSS, DLR-QMS, including an audit and alert system, documented assembly and integration, among other measures.
- A risk management system, that is risk analysis and risk evaluation according to MIL-STD-882D and ECSS.

For the design lifetime of one year the selection of the electrical, electronic and electromechanical (EEE) parts followed a mixed strategy. In general, all parts used for the satellite implementation have the "industrial-type" or commercial off-the-shelf (COTS) quality level, except for components that were identified as weak points from the reliability point of view by special analysis (like FMECA etc.). In these cases parts qualified according to MIL specifications were used. Using this approach the space segment costs and the risk for the one-year design lifetime could be kept low. The BIRD mission was developed, built, tested and placed into orbit within five years for a total cost of about 15 million euros. The designed lifetime of one year has been exceeded by several years of successful operation.

8.5.2.3 The Launch Element

In accordance with the design-to-cost philosophy the satellite was planned for launch as a secondary payload. This required the satellite design to be suitable for different LEO orbits and adaptable to different launcher requirements. Even though the payload with the cooled infrared sensors had high power consumption, the spacecraft had to be small and light. The microsatellite had to be compatible with the launch adapters of different launchers like Kosmos, Dnepr, Zenit, PSLV, Ariane and others. The dimensions of the satellite had to be less than $500 \times 500 \times 800\,\text{mm}^3$ and the total mass less than 100 kg. Keeping the footprint of the satellite within the required dimensions was usually more important than maintaining the orbital altitude of the satellite, within limits. Another important launch requirement was easy mechanical adaptation to different launch platforms from the design and construction points of view. The qualification of the satellite including the adapter plates, if applicable, had to apply to different launchers.

8.5.2.4 Orbit and Constellation

The objectives of the BIRD mission can be achieved by one single satellite in a low Earth orbit. According to the design-to-cost philosophy, it is not possible to define the orbit yielding the maximum scientific output or best operational performance. But it was important to define the crucial orbit parameters and limits in order to identify one or more launch options in the planned launch year of 2001. The primary orbit parameters for a nearly circular orbit are the average altitude and the inclination of the orbit. For the BIRD mission, orbit altitude is determined by:

- The design lifetime in orbit and the descent rate due to atmospheric drag (lower limit of altitude)
- The required geometric and radiometric resolution in combination with the optics design (upper limit of altitude).

Also, the following limits of a nearly circular orbit were defined:

- Desired orbit altitude: 450–650 km
- Maximum orbit altitude: 850 km.

The inclination influences different important mission parameters like area and time coverage, repeat time, illumination conditions of the target (local time), regression of the ascending node, ground station contacts and other factors. For the BIRD mission a fixed orbit inclination was not required.

The upper limit of inclination is the desired Sun-synchronous inclination and the lower limit is determined by the requirement to cover the German ground segment. The required orbit inclination i is defined as

$$53° \leq i \leq 99°.$$

All other orbit parameters (for instance, equatorial crossing time, right ascension of the ascending node, eccentricity, etc.) cannot be defined by a secondary payload passenger.

8.5.2.5 The Space Element

The BIRD microsatellite is the space element of the mission. It consists of the payload and the spacecraft bus. Some basic parameters are summarized in Table 8.5.2.

The payload of the space element results from the geometric, radiometric, spectral and other requirements derived from the mission objectives. Some basic requirements are summarized in Table 8.1.1 in Section 8.1.4. The main equipment of the BIRD payload consists of:

Table 8.5.2: *Basic parameters of the BIRD satellite.*

Parameter	Value
Satellite mass (total)	92 kg
Payload mass	30 kg
Power consumption average/peak	60 W/210 W
Stabilization	Three-axis stabilization
Communication	S-band, 2 Mbit/s
On-board mass memory	2 × 1 Gbit
Design lifetime in orbit	1 year

- A bispectral infrared sensor system for the detection of hotspots (heat sources)
- A wide angle optoelectronic stereo scanner (WAOSS-B)
- A payload data handling system with 1 Gbit mass memory
- An artificial neuronal network classification experiment.

Some basic parameters of the payload sensor system are given in Table 8.5.3. Several requirements for the spacecraft bus result from the payload, the operational concept and the sensor system parameters. The BIRD satellite design is characterized by a compact architecture consisting of three segments: the payload segment, the electronics segment and the service segment. The payload segment includes the scientific instruments (see Table 8.5.3) and some equipment of the satellite attitude sensor system on a separate platform. This kind of separation allows mechanical–dynamic, thermal and manufacturing decoupling from the spacecraft bus and thus time-optimized parallel assembly of the payload segment and the spacecraft bus. The payload platform, consisting of a multiple sandwich construction of carbon fiber composites, has high thermal–optical stability and forms a simple interface with the satellite bus.

The payload segment with its scientific instruments, star sensors and magnetometer is an autonomous unit. The payload segment is geometrically and radiometrically calibrated in the calibration lab by means of extensive optical ground support equipment (OGSE). Following the integration and test plan, the payload platform is integrated with the

Table 8.5.3: *Overview of the BIRD sensor system.*

	WAOSS-B	MIR	TIR
Wavelength	600–670 nm 840–900 nm	3.4–4.2 µm	8.5–9.3 µm
Focal length	21.65 mm	46.39 mm	46.39 mm
Field of view	50°	19°	19°
f-number	2.8	2.0	2.0
Detector	CCD arrays	CdHgTe arrays	CdHgTe arrays
Detector cooling	Passive, 20 °C	Stirling, 80 K	Stirling, 80 K
Pixel size (µm × µm)	7 × 7	30 × 30	30 × 30
Pixel number	3 × 2884, in use	2 × 512, staggered	2 × 512, staggered
Quantization	11 bit	14 bit	14 bit
Ground pixel size[a]	185 m	370 m	370 m
GSD[a]	185 m	185 m	185 m
Swath width[a]	533 km	190 km	190 km
Power consumption	18 W	42 W +90 W cooling	
Mass	8.4 kg	8.7 kg camera head +5.8 kg electronics	

[a] Orbit altitude = 572 km.

WAOSS-B, Wide Angle Optoelectronic Stereo Scanner; MIR, Medium-wave Infrared Sensor; TIR, Thermal Infrared Sensor; GSD, Ground Sampling Distance.

complete, separately assembled and tested spacecraft bus as the next step. The payload segment and the satellite bus main body together form the cube-shaped satellite body as the primary structure in a compact assembly. One body-fixed and two foldable solar arrays are mounted onto the primary structure.

The mission constraints require a microsatellite to meet the mission objectives. That means a satellite with a total mass of about 100 kg is launched together with a primary payload into orbit. The mission objectives and the payload require many high-performance components from the satellite bus. They are established with many innovative technological solutions

within a strict cost limit. Essential characteristics of the BIRD satellite bus are:

- Compact microsatellite structure with high mechanical stability and stiffness and adaptability to different launcher platforms.
- Envelope qualification for several launchers (Cosmos, PSLV, Dnepr, and others).
- Cube shape in launch configuration with dimensions of about 620 × 620 × 550 mm³.
- Mass ratio of bus to payload = 64 kg:30 kg.
- High peak power of 200 W for 10–20 min, and average power 60 W.
- Passive thermal control system with radiators, heat pipes, MLI, sensors and a few emergency heaters.
- Newly developed high-performance spacecraft bus computer with integrated latchup protection and an error-detection and correction system.
- Three-axis stabilization by an attitude control system with newly developed precision reaction wheels and newly developed star sensors.
- On-board determination of the spacecraft's position and velocity based on GPS data reception and an on-board orbit model (10 m accuracy has been demonstrated).
- S-band communication with high bit rate (2.2 Mbit/s) and low bit rate for command and telemetry transmission.
- Redundant on-board payload data handling system with 1 Gbit mass memory.

The constructive solution is a **compact satellite** in a cube shape with three segments (see Figure 8.5.1). The service segment contains satellite-specific equipment like batteries, reaction wheels and an inertial measurement unit (IMU). The electronics segment comprises the spacecraft bus computer, most of the electronic units of the satellite bus, and the redundant payload data handling system. As mentioned above, the payload segment contains the exchangeable payload platform with the instrumentation. The term **compact satellite** expresses the high integration density of all equipment and components within the primary structure, independent of their subsystem assignment. This is a very common architecture for microsatellites. In contrast to the compact architecture, a satellite can

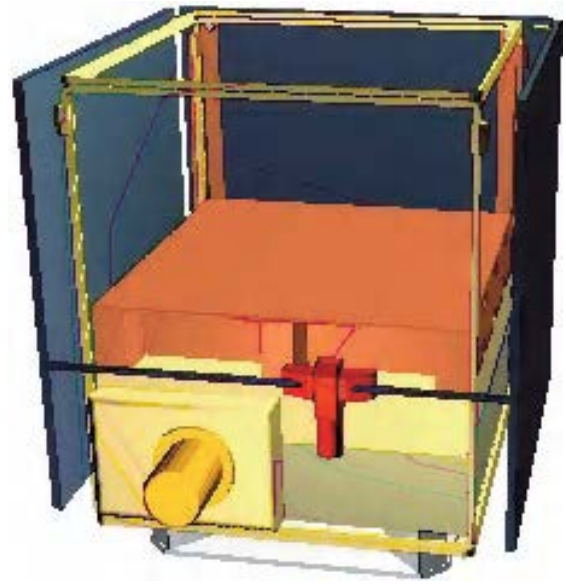


Figure 8.5.1: *BIRD satellite bus as a compact satellite with cubic shape, consisting of three segments: exchangeable payload segment (free space), electronics segment (red), service segment (yellow body) (Source: DLR).*

be designed in a **box architecture** characterized by an arrangement of satellite equipment in functional box units, for instance an electronics box, a telemetry box, a payload data handling box, and so on.

The BIRD satellite bus includes several advanced technologies that are space flight proven with this mission. The satellite, including the integrated payload platform, is depicted in Figure 8.5.2 in flight configuration. An overview of the characteristic equipment of the subsystems is given in Table 8.5.4.

From the very beginning the BIRD satellite worked very successfully and demonstrated the performance parameters of the advanced technologies under space conditions. Some of them were implemented in the equipment of the attitude control system because of their importance for remote sensing satellites.

The equipment for the attitude control and navigation system is depicted in Figure 8.5.3.

The requirements for the attitude control system are determined essentially by the spacecraft modes and the payload operations. They are presented in Table 8.5.5 for the Earth observation mode of BIRD.

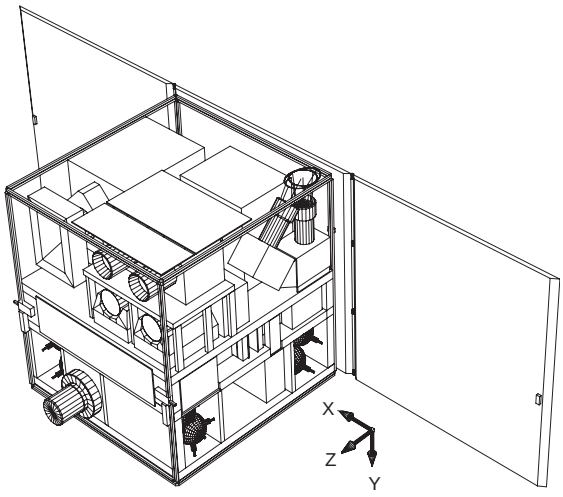


Figure 8.5.2: *BIRD satellite in flight configuration (Source: DLR).*

Table 8.5.4: *Overview of the BIRD satellite bus.*

Subsystem	Equipment
Attitude control subsystem (ACS)	Two star sensors (accuracy 10"), three-axis gyroscope (accuracy 2.7"), magnetometer, GPS receiver, four reaction wheels (max. momentum > 0.2 N m s), three magnetic coils (max. 3 A m²); ACS computer = on-board computer
Board computer (OBDH)	Processor 80C517 (Power PC), 512 kbyte dual-port RAM
Tracking, telemetry and command subsystem	High- and low-gain antenna for S-band, antenna switches, two S-band receivers, two S-band transmitters (BPSK, max. 5 W RF power)
Structure and mechanisms	Base plate, electronics compartments, frames, fastening elements, deployment mechanism, eject mechanism
Electrical power subsystem	Power generation: three solar panels with Si-high-η cells; power storage, eight NiH_2 cells with 12 Ah; shunt charge regulator, direct energy transfer, power distribution: unregulated 20 V bus, harness
Thermal control subsystem	MLI, infrared system radiator, satellite radiator, heat pipes, heat conductors

The following attitude control equipment was newly developed: the attitude control algorithms and software, the star sensors, the magnetic coils, the magnetic coil control, the reaction wheels including internal hardware and software, and the board computer hosting the attitude control software. To reduce the technical risk connected with new developments, the following measures were adopted:

- Extensive development tests and qualification tests on the equipment level
- Redundant implementation of all critical units or equipment.

Redundancy can also be functional redundancy. For instance, the two star sensors and the gyroscope establish a 2-of-3 redundant attitude sensor system. The three-axis magnetometer supplements this redundant system as a solution for a further emergency case, but with very limited performance. The precision reaction wheels are implemented as a hot redundant system: although three reaction wheels are enough for the required precision, BIRD has four reaction wheels running in the nominal case. The magnetic coils for the desaturation of the reaction wheels are implemented in cold redundancy. In an emergency attitude control can be accomplished by magnetic coils as the only actuators, but with a very limited performance and only to a certain extent. The navigation package contains an on-board orbit propagator developed by DLR-GSOC and fitted by measurement points from the GPS receiver. If the GPS receiver fails, the supporting points for orbit calculation can be uploaded from the ground. Of course, the positional accuracy is decreased.

The attitude control software runs on the spacecraft bus computer (= board computer), which is connected to a second computer in a master–slave configuration. A failure of the master is detected by the slave, which monitors the master permanently. In case of failure the functions switch and the slave takes over all the functions of the master.

8.5.2.6 Communication Architecture

Communication between the BIRD satellite and the national ground segment during nominal operations

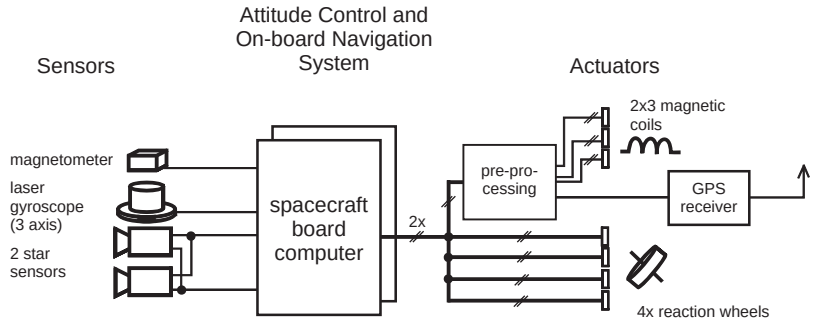


Figure 8.5.3: *Block diagram of the attitude control system of BIRD.*

is defined as point-to-point architecture as a basic principle. The BIRD satellite has an internal data storage capacity for a measurement sequence of 10 min and provides the option of store-and-forward transmission. That means the satellite can take image data from any point on the Earth, store it and then dump the data down on the next pass over the German ground station. Simultaneous image acquisition and downlinking over Europe is also possible. Figure 8.1.2 in Section 8.1 shows the mission architecture, including the communication architecture. Command of the satellite is carried out exclusively by DLR-GSOC, with the mission control center at Oberpfaffenhofen and the ground station at Weilheim. Data reception during nominal operations is accomplished via the ground stations at Neustrelitz and Weilheim. In the Neustrelitz data receiving station the received data is decommutated and the telemetry data is transmitted to Oberpfaffenhofen for processing and evaluation. The scientific data is processed systematically and archived in Neustrelitz.

8.5.2.7 Satellite Ground Stations

As already described, routine operations are performed via the DLR satellite ground stations in Weilheim (primary ground station) and Neustrelitz (science ground station). Additionally, an experimental ground station at Berlin-Adlershof has been established. This ground station demonstrates the capability of immediate reception of BIRD data by a regional or local user somewhere in the world. It consists of a small controllable antenna dish (diameter 2.40 m) and appropriate receiving and processing units. It is tailored to the demands of regional users wanting to evaluate the satellite data without delay. The experimental ground station consists of only a few units that can be transported without great effort to any receiving point on the Earth and then assembled and operated there as a receiving station for the end user of the data, for instance for firefighting coordination.

Additionally, partner ground stations can be added as part of the point-to-point communication architecture. This was done temporarily with ESA and the Argentinean space organization CONAE. They received without delay all BIRD data for the local area and processed the data into final data products within 20 minutes.

For the most critical phase, the initial acquisition of data after launch, additional ground stations at Kiruna (ESA) and Fairbanks, Alaska (Prioranet), were involved. So in the launch and early orbit phase all satellite data could be received every 1½ hours. After the first day this active support was changed to standby support for emergency cases.

8.5.2.8 Mission Control Center and Mission Operations

The mission control center for the BIRD mission at DLR in Oberpfaffenhofen incorporates the satellite control center and the payload control centre – a common approach for small-satellite missions. The users address their remote sensing requests to the science team or directly via the Internet into the mission planning tool. The remote sensing requests from the different users are coordinated by the science team. Technical constraints such as power status, free data memory capacity or next orbit for the target pass are adjusted by the mission control center.

Mission operations for the BIRD satellite are complex because of the number of limitations. The satellite cannot measure continuously with all instruments but for 10 minutes in one orbit only and also not in each orbit. The reason is the unusual (for microsatellites) high peak power consumption of 200 W in the Earth observation mode with simultaneous data dumps. Generally microsatellites generate an average power of 20–60 W. But BIRD is an experimental satellite without a continuous duty cycle, so the measurement time is limited to 10 min. The associated swath length on the ground of approx. 4000 km is completely sufficient for experimental purposes. The energy for the 20 min cooling-down time and 10 min measurement time is taken primarily from the batteries. After this period the spacecraft slews itself so that the solar arrays point perpendicular to the Sun to recharge the batteries. The operational modes of the satellite are depicted in Figure 8.5.4. The basic mode is the Sun-pointing fixed mode (SPF in Figure 8.5.4) with the solar arrays perpendicular to the Sun vector and the satellite moving inertially fixed. The same attitude is assumed by the satellite in safe mode and in the auto-acquisition mode (AAM), with the satellite detecting the Sun autonomously and aligning the solar arrays. The roughly nadir-pointing direction of the payload segments and the high-gain antenna are used for ground contact and high data rate transmission.

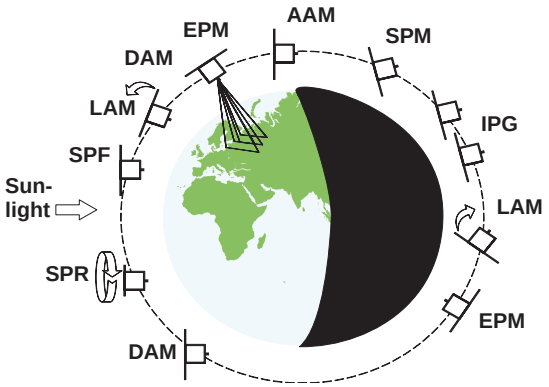


Figure 8.5.4: *Satellite modes of BIRD: AAM, Auto-Acquisition Mode; DAM, Damping Mode; LAM, Large-Angle Maneuver; SPF, Sun-Pointing Fix; SPR, Sun-Pointing Rotate; EPM, Earth-Pointing Mode; IPG, Inertial Pointing Mode; SPM, Suspend Mode.*

The Earth-pointing mode (EPM) requires precise pointing accuracy, as described in Table 8.5.5. Besides these main operational modes, the satellite can carry out an on-board data processing task as a technological experiment. The processing results are high-level data products of the type so far produced only on the ground. Earth observation images can be recorded during nighttime, too. From time to time the satellite points into deep space or to the Moon for radiometric calibration purposes (IPG, Inertial Pointing Mode). Changes of the attitude modes are accomplished by so-called large-angle maneuvers (LAMs). Damping of the tumbling movement after separation from the upper stage is executed in the damping mode (DAM). In the suspend mode (SPM) almost all attitude control equipment is switched off.

8.5.2.9 Data Processing, Archiving and Distribution

The processing, archiving and distribution of BIRD payload data are accomplished immediately after reception by the DLR ground station at Neustrelitz. The science team provides the algorithms for systematic data processing. These are implemented, tested and

Table 8.5.5: *Requirements for navigation and attitude control of BIRD in the Earth observation mode.*

Category	Requirement
Orbit determination	
Position in orbit, each axis	100 m
Time	20 ms
Attitude determination	
Measurement precision	0.1 mrad ≈ 0.2 arc min
Line of sight of payload and high-gain antenna	Earth, nadir, space, Moon
Attitude control	
Pointing accuracy, each axis	±2 mrad ≈ 7 arc min (3σ)
Pointing stability (jitter)	5 arc min/s
Slew rate	1.0°/s
Drift rate	1°/h, 5° max.
Slew range, each axis	±180°

executed after each data reception in the processing center. The raw data and the data products are stored in an archive system and produced on demand. The data processing level of the BIRD mission is in accord with the international definitions of the Committee on Earth Observation Satellites (CEOS). These are explained in [8.5.2] and listed in Table 8.5.6. The scientific users receive data processed at each level according to their request. Especially for semi-operational services for vegetation fire management or for field experiments, level 2 data (fire data products) is very useful.

8.5.2.10 Users

The users of the BIRD mission are scientists and experts in different fields. A large user group consists of fire ecologists and international scientists dedicated to remote sensing of vegetation fires, such as the group at the Global Fire Monitoring Centre in Freiburg/Breisgau, Germany. But volcanologists and geologists investigating coal seam fires are also users, as are to a certain extent administrators and fire department managers. Another user group consists of engineering experts in infrared remote sensing technologies and microsatellite engineers. These users belong to national research organizations (DLR, Fraunhofer Gesellschaft), universities, ESA, and several small and medium-sized companies which contribute their own resources toward BIRD's mission success.

Table 8.5.6: *Levels of data products in the BIRD mission according to [8.5.2].*

Level	Data type
Level 0	Unprocessed instrument/payload data in combination with spacecraft data, communication frames/headers removed
Level 1A	Time-organized single sensor raw data with an appendix including: • instrument housekeeping data • radiometric and geometric calibration coefficients • georeferencing parameters (ephemeris data) • other ancillary information
Level 1B	Radiometrically and geometrically processed level 1A data in sensor units (radiometrically and geometrically calibrated sensor data)
Level 2	Interpreted geophysical parameters (hotspot temperatures, hotspot extension, vegetation indices, cloud parameters)

8.5.3 System Integration and Verification

System integration of the flight model of the BIRD satellite involves as the main step the integration of the payload platform comprising all the integrated and calibrated instruments with the spacecraft bus, followed by the installation of the multilayer insulation. The next step is the assembly of the secondary structure, namely the three solar arrays. For system integration a clean room of class 200 000 is sufficient, because no dust-sensitive surfaces are endangered. Nevertheless, a common clean-room regime is kept, like the wearing of special clothes, and access and behavior regulations. But costly and technically elaborate forced air circulation and cleaning is not used.

A crucial point to comply with fixed cost limitations is the selection of the model philosophy. For the BIRD space element a hybrid model philosophy was defined to meet the requirements for a high-quality-level spacecraft under the condition of making extensive use of innovative solutions. On the equipment level, there are different technology maturity levels and corresponding individual model philosophies. The (preliminary) status of the BIRD model philosophy as of January 28, 2000 was defined in the "hardware matrix" document as shown in Table 8.5.7 as an example [8.5.3]

According to the hybrid model philosophy, the following models were built on the system level.

Mass Model

The **mass model** (MM) was built for the qualification of a separation system developed in-house but not used later. However, the mass model was delivered to the launch pad long before delivery of the flight model, just in case the flight model was not available during the launch preparation phase.

Table 8.5.7: *Hardware matrix of the BIRD satellite [8.5.3].*

	STM	EM	PFM
Lower structure box	(F)	D?	F
Upper structure box	(F)	–**	F
Launch adapter	F	–	F
Solar arrays	D*	–	F
MLI	TM	–	F
Harness	D*	EEM	F
ACS			
Reaction wheels	TM	4 × EQM	4 F
Magnetic coils	2 × 3F	–	2 × 3F
Star sensor	TM	1EQM	2 F
Gyroscope	TM	F	F
TT&C	2 TM	1 EEM	2 F
Satellite computer	TM	EM	FM
Power			
Battery	TM	EEM	8 FM
PCU	TM	EM	
PDU	TM	EM	
Thermal control	TM	–	F
Command unit	TM	EM	F
WAOSS TM	QM	F	
Infrared system	TM	–	F
PDH	TM	EM	F
Payload platform	F	–	F

* No FM representative.
** Side panels of PFM.
F, Flight Standard; D, Dummy (mechanical); TM, Thermal Model; EEM, Electrical Engineering Model; EQM, Engineering Qualification Model; STM, Structure and Thermal Model; EM, Engineering Model; PFM, Prototype Flight Model.

Structure and Thermal Model

The **structure and thermal model** (STM) of the BIRD satellite was built to qualify the design of the structure for the launch loads according to the specifications, to verify the mechanisms, and to verify and improve the thermal control subsystem and the thermal–mathematical model of the satellite. The integration of this model was finished in March 1999 and qualification tests using the vibration test facilities (shaker) and the thermal vacuum chamber (see Figure 8.5.5) were completed in May 1999. In the year 2000 the STM was also used for vibration tests at the qualification level for the upper stage of the PSLV launcher provided by India.

Engineering Model

After termination of all test activities with the structure and thermal model it was partly decommissioned and rebuilt as the engineering model of the satellite for cost reasons. The **engineering model** (EM) consists of all essential electrical, electronic and electromechanical equipment of the satellite in flight configuration but not in flight model quality level. It is used to test and verify all essential electrical, electronic and electromechanical functions and the proper functioning of the interaction between hardware and software. In order to build a model with an arrangement of all units according to the flight configuration, the structural elements of the STM are used to build the engineering model (see Figure 8.5.6). For the payload no engineering model was available, only the flight model because of the costs. The qualification of the payload equipment was carried out individually on the equipment level.

Prototype Flight Model

The **prototype flight model** (PFM) of the satellite is actually used to carry out the mission. The model was tested at the qualification level but with reduced test



Figure 8.5.5: *Structure and thermal model (STM) of the BIRD satellite in front of the space simulation chamber ready for the thermal vacuum tests (Source: DLR).*

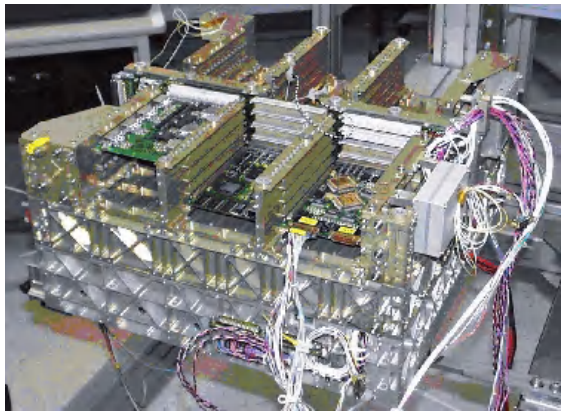


Figure 8.5.6: *Engineering model of the BIRD satellite bus (Source: DLR).*

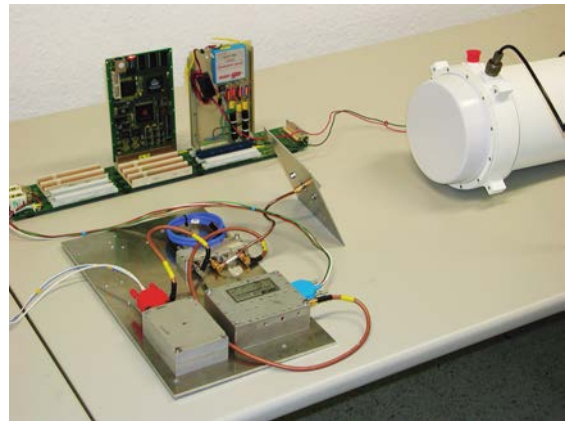


Figure 8.5.7: *Suitcase model of the BIRD satellite and antenna feed of the S-band ground station. The transmitter, receiver, antenna switch and antenna model are shown at the front and the spacecraft bus computer with modulator, switch and power supply are shown at the back. The antenna feed of the ground station can be seen on the right (Source: DLR).*

durations. The subsequent functional test completed the PFM.

Further models

For certain aspects of quality assurance and for verification of the system, further special models were built.

A dedicated experimental infrared airborne system (**airborne simulator**) was built to verify the advanced infrared technology and special data analysis methods like the Dozier method. With this development model important knowledge was gained before launch and also during the routine mission operations of the BIRD satellite. The spaceborne remote sensing data could be verified by the airborne system.

The lines of sight of all three BIRD instruments (WAOSS-B and two infrared instruments) have to maintain their directions very precisely with respect to each other (less then 1 mrad deviation) despite thermal conditions and changes. To verify the stability of the lines of sight with respect to each other an STM of the payload was built and tested in a space simulation chamber. Changing orbit and operational conditions were simulated during an entire week. This payload model was later integrated in the STM of the satellite.

Another necessary model was the so-called **suitcase model**. It is required for compatibility tests with all planned ground stations. For transportation to all the ground stations it is required to be small (Figure 8.5.7), but the modulation, frequency and other RF properties must be identical to the flight model.

The correct functioning of the attitude control for the BIRD satellite is a precondition for survival in space. For this reason verification of the attitude control system was accomplished in two steps. In the first step all attitude control software was verified by a software simulation package. In the second step a complete attitude control model of the satellite with all its equipment was built and mounted on an air bearing table. This experimental model was used for **hardware-in-the-loop simulation** of the attitude control system including the flight software. The following hardware units were involved in the simulation: receiver, demodulator, spacecraft bus computer, three-axis gyroscope, four reaction wheels, 2 × 3 magnetic coils, Sun sensors, magnetometer. Experimental verification of the following functions was performed:

- Three-axis stabilization.
- Finding the Sun vector and slewing the solar arrays to the Sun.
- Holding the Sun direction (control loop).
- Performance of defined slew maneuver.
- Performance of desaturation of the reaction wheels by means of magnetic coils.

The use of the different models on the system level for system verification is presented in Table 8.5.8.

Figure 8.5.8 shows an overview of the hybrid model philosophy of the BIRD mission. The relationships between the models on the three levels of equipment, subsystem and satellite are represented graphically. The rectangles are the models and the ellipses describe the activities. After completion of the qualification of the structure and the thermal control subsystem the STM was decommissioned and its equipment and components were partly reused to build the EM to save costs. In addition, the flight model was built. Because of this procedure the approach is called the 2½ model philosophy. It assures maximum qualification on the system level at moderate total costs for a microsatellite with a high degree of innovative equipment. Accordingly, for comparable microsatellite projects it is recommended that the potential risks and the system drivers for qualification of the satellite are identified in order to derive a tailored and cost-effective model philosophy.

Table 8.5.8: *BIRD verification on the system level.*

	STM	EM	PFM
Electrical function tests, performance tests	×	×	×
Functional instrument tests	×		×
Payload data system tests			×
Physical properties	×		×
Sweep sine test	×		×
Random vibration test	×		×
Sine vibration test	×		×
Thermal cycle test	×		×
Thermal vacuum test	×		×
Thermal balance test	×		×
Solar panel deployment test	×		×
Match mate test			×
Mission operations tests	×	×	×
Software upload test		×	×
End-to-end system test			×

Bibliography

[8.5.1] IAA Position Papers, Subcommittee on Small Satellites: The Case for Small Satellites. *Acta Astronaut.*, **31**, 103–144, 1993.

[8.5.2] Brieß, K., Lorenz, E. Systematic Image Processing of the Small Satellite Mission BIRD. Proceedings of the 49th International Scientific Colloquium, Ilmenau, September 27–30, 2004.

[8.5.3] BIRD. Technical Note, January 28, 2000.

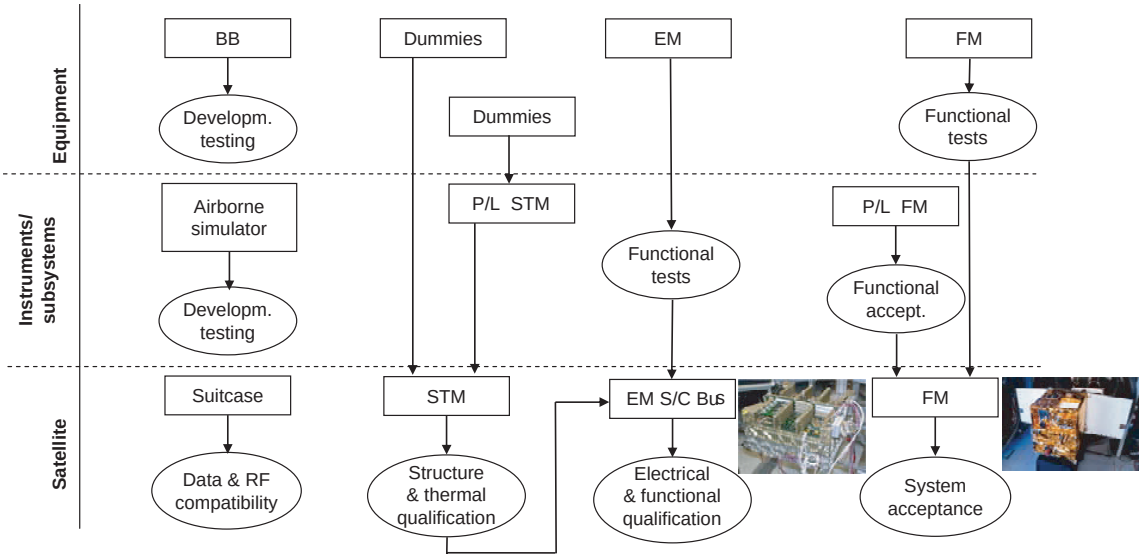


Figure 8.5.8: *Hybrid model philosophy of the microsatellite BIRD: BB, development models on the equipment level (Breadboards); EM, Engineering Model; FM, Flight Model; STM, Structure and Thermal Model; P/L, Payload, S/C, Spacecraft; RF, Radiofrequency.*

8.6 Galileo Satellites

Hanspeter Kuhlen

8.6.1 System Requirements

As a further example, the following sections describe the satellites of the European radio navigation satellite service (RNSS) named Galileo, a system for the precise **three-dimensional determination of local position** (longitude, latitude, altitude) and accurate time provision (UTC and clock seconds) anywhere on the Earth's surface as well as up to altitudes of several thousand kilometers. The latter feature enables space vehicles to determine easily and precisely their own orbital positions and instantaneous attitude by using three antennas located on the edges of the vehicle's structure.

The complete Galileo system comprises **27 active satellites** of which nine plus one additional spare satellite travel in three different orbital planes, each inclined by 56° to the equatorial plane. The ascending nodes (right ascension of the ascending node (RAAN)) of each of the three planes are separated by 120°, leading to the best possible equal distribution of spacecraft around the globe. The spare satellite in each orbital plane is in "hot standby" mode with only the radiofrequency (RF) power stages switched off. It can replace a faulty satellite within a few hours or days depending on the position of the defective satellite. At an orbital altitude of about 23 620 km above the Earth, the satellites travel in a sometimes-called **medium Earth orbit** (MEO) in a zone which imposes particularly high doses of space radiation on the satellites. The Galileo constellation will enable a global accuracy of 4 m laterally (2σ value) and 8 m vertically (2σ value) without regional or local augmentation support for more than 95% of the time. A sophisticated terrestrial performance measurement network consisting of more than 30 sensor stations determines for safety-critical navigation users (e.g., aviation) the so-called "integrity" of the instantaneously achievable accuracy. Galileo safety-of-life receivers on-board aviation vehicles will receive an integrity warning message within 6 s of the detection of conditions exceeding a specified threshold, for example, 20 m vertically and 12 m laterally.

In addition to the position–navigation–time (PNT) services with and without integrity information, the system also provides a global **data dissemination service**, the so-called commercial service (CS). The CS transmits navigation-related encrypted information with a net bit rate of 500 bit/s. Furthermore, there is a special payload dedicated to permanently monitoring the international distress call frequency on 406 MHz for calls originating from **aviation and maritime search and rescue** (SAR) distress beacons. This payload operates within the framework of intergovernmental COSPAS-SARSAT agreements. Galileo SAR support not only enables active acknowledgment of a received distress call through a return channel, but also performs active localization by the radio triangulation of legacy beacons without integrated RNSS receivers. All Galileo satellites in the in-orbit validation (IOV) and full deployment (FOC) phases are designed for an in-orbit lifetime of 12 years. Although this complies with the state of the art of telecommunications satellites, it represents a major challenge for the long-term operation of on-board (ultrastable oscillators) atomic clocks on navigation satellites. All navigation signals and RF link parameters are designed to be **compatible with the transmissions of other RNSS systems**. The unrestricted signals of the Galileo OS (Open Service) and SoL (Safety-of-Life) services are also designed to be interoperable with the co-frequency signals of the US GPS system. This means that dual-system receivers can constructively process the signals of both systems into a valid navigation and time solution. Figure 8.6.1 shows the **navigation signals** E5, E6 and E1 transmitted permanently from each of the Galileo satellites in the respective RNSS-allocated frequency bands with corresponding bandwidths. Note that the E1 signal was called L1 in earlier program phases, creating confusion with the GPS-L1 since the signal structure in Galileo differs significantly from the GPS-L1.

Each transmitted signal consists of a complex multiplex of in-phase and quadrature phase signal components carrying two I and Q (E5) and three A, B and C (E6, E1) orthogonal channels. All signals have a so-called pilot channel in common, intentionally carrying no data messages and thus optimizing solely the ranging performance, in particular under

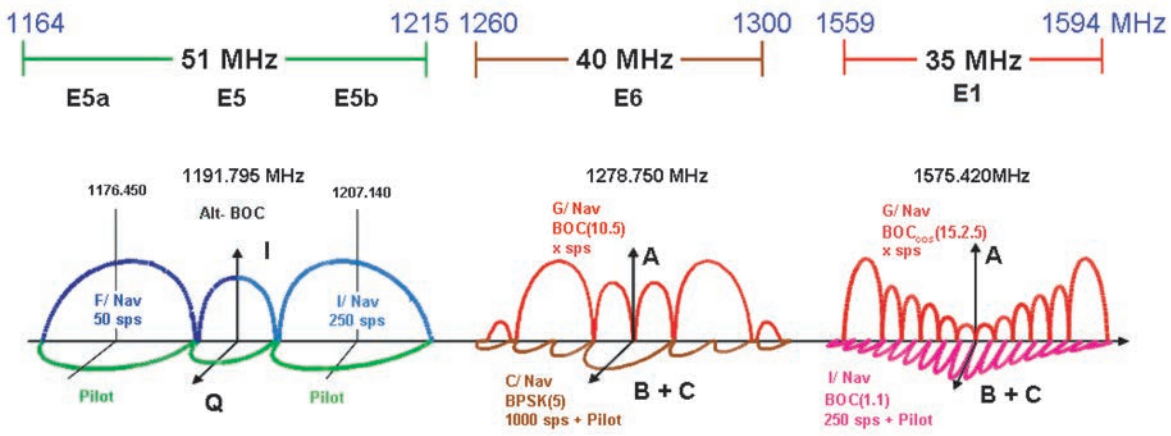


Figure 8.6.1: *Typical signal and frequency plan for Galileo (Source: EC/ESA).*

critical signal propagation conditions such as indoor reception. This results in up to 10 signal components being processed in user terminals either as stand-alone or in any reasonable signal combination. This makes possible a wide variety of low-cost to high-end user receivers for a wide range of target groups and applications. For instance, the use of dual-frequency receivers can increase the timely availability of a high accuracy of less than 1 m because it allows the receiver to calculate a sophisticated ionospheric model to compensate for the impact of the variance in the signal arrival time due to varying propagation conditions in the ionosphere.

The **fine structures of the signals** are optimized to meet the demands of the three main target user groups. These are:

- **Private and commercial users**
- **Safety of Life (SoL) users** in the aeronautical, railway and maritime safety-critical areas
- **Governmental and security-sensitive user** groups (public regulated service).

8.6.2 Design Driver and Design Process

All design requirements represent the results of many comprehensive and dedicated studies aggregated in a requirements database from which all interface, subsystem and equipment specifications are derived and defined. The set of specifications forms the basis for the entire development and flight qualification process of the Galileo satellites. Each satellite contains a support module, also called the **satellite bus** or **platform**, and a **payload** module specifically tailored to the Galileo mission. The specific requirements for the payload largely drive the framework conditions for the satellite design, such as the accommodation of "cold" and "hot" equipment, their power supply, as well as the RF cable connections to the antenna farm. This in turn drives the dimensioning of the payload-supporting subsystems, particularly the power supply designed for the specified end-of-life performance after 12 years, and the thermal subsystem that drives the satellite volume and mass.

The total number of flight units for the Galileo system represents a large number, 30 plus ground spares for a space business accustomed to producing large satellites mostly in single units. In principle, satellite navigation systems can only provide good 3D positioning if at least four satellites with large angular separation are in view of an observer's position. Design changes or new features only become perceivable when all the satellites of a constellation provide the same degree of service. Thus, the large number and the fact that all satellites have to be identical to provide equal performance lead to the exact duplication of satellites regarding design, performance and

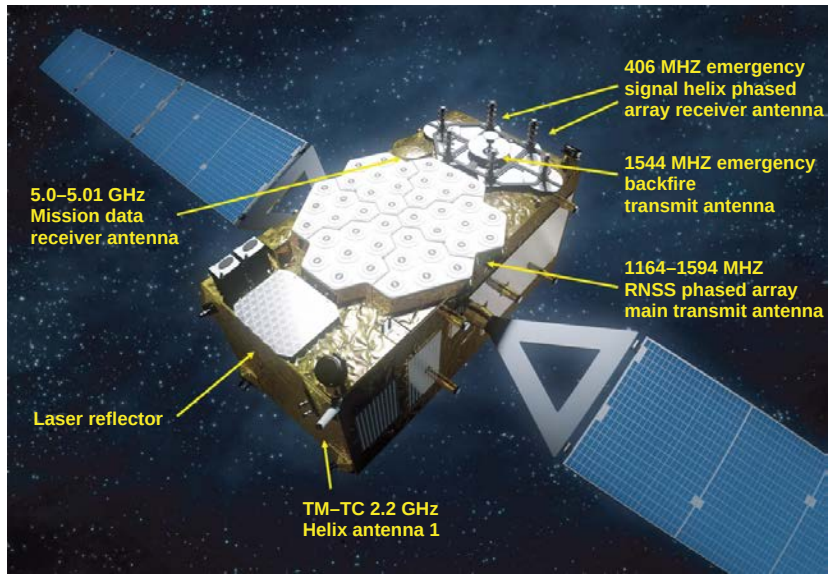


Figure 8.6.2: *Galileo satellite in flight configuration (Source: ESNIS).*

fuel loading (propulsion performance). With higher quantities of satellites, it becomes more important to consider an optimized design for manufacturing to achieve cost advantages during satellite production and testing. An impression of a "ready to fly" Galileo satellite with a view of its Earth-pointing antenna farm is shown in Figure 8.6.2. To ensure consistent orientation during design, development and documentation, an **orthogonal (Cartesian) coordinate** system has been defined with the reference point in one corner of the satellite, as shown in Figure 8.6.3. This coordinate definition yields a positive number for each reference to panel positions, interfaces or any other point on the spacecraft. Furthermore, the origin of the coordinate system is located at a point with good accessibility on the surface of the spacecraft. It exists physically at a very early stage of the integration process, providing clear references for all follow-up alignments during integration. The positive *x*-axis of the Galileo satellites points in the flight direction while the positive *z*-axis points toward the center of the Earth to optimize the direction of the boresight gain of the main mission antenna toward the Earth, and the solar arrays turn around the *y*-axis to maintain Sun pointing. The Galileo satellites' yaw around the *z*-axis (yaw-axis) in the course of one orbit describes a circle with the *x*-axis. However, before the positive *x*-panel enters into

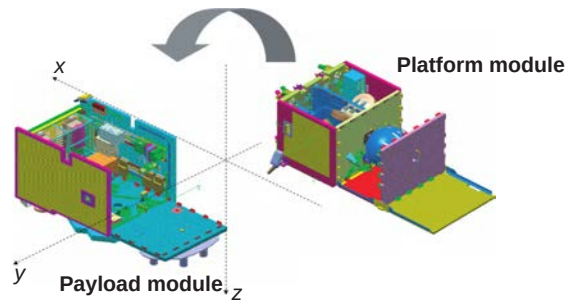


Figure 8.6.3: *Satellite platform and payload modules with coordinate system (Source: ESNIS).*

sunlight, the attitude control system forces the satellite to maneuver around the *z*-axis and this attitude correction is repeated for each orbit.

The *x*-panel never sees the Sun because of this maneuver; therefore it can **radiate all dissipated heat** from the spacecraft into cold space. With the *z*-axis pointing of the main mission antenna toward the center of the Earth (nadir pointing), the antenna provides a nearly constant power flux density on the visible surface of the Earth because of its **isoflux characteristic**. This special characteristic assures that more RF power is radiated effectively toward the edge of coverage rather than to the subsatellite point.

Connected to the *y*-walls are two solar arrays each providing **850 W of end-of-life DC power**; the rotation around the *y*-axis maintains a Sun-pointing mode controlled by the on-board computer (integrated control and data unit). During the relatively long launch implementation process of direct injection into the final orbit, the solar arrays are folded into the *y*-walls and do not produce any power. As soon as the solar arrays are deployed, the *y*-walls offer further capacity for heat dissipation into space. Dissipation areas (panels/walls) toward cold space are key elements for sustaining a comfortable operating temperature inside the satellites since high-power devices can generate considerable amounts of heat. Because the vacuum in space prevents the transportation of heat through convection, any source of heat is thermally connected to heat-conducting surfaces and/or heat pipes. The right mix of active and passive thermal elements ensures **balanced thermal conditions** under all specified satellite operating conditions. The attitude and orbit control subsystem maintains the **three-axis stabilization** by orientation of the Earth- (*z*) and Sun-pointing (*y*) axes based on real-time information provided by appropriately mounted Sun and Earth sensors. Intentional changes of attitude as well as compensations for orbital disturbances can be performed by monopropellant thrusters, momentum wheels and magnetic torque rods that are geometrically mounted such that the on-board computer can command any required attitude. These control processes can be either commanded through ground control or performed autonomously. In summary, it can be concluded that the main design drivers are:

1. The **primary power** required at the end-of-life of the spacecraft.
2. The **launch mass at the beginning of a mission**, including fuel for all maneuvers throughout the satellite's lifetime until final transfer to a graveyard orbit.
3. The **overall volume** of the satellite with solar arrays stowed.

The fairing geometries of the four **launch systems** that are to be considered as launch alternatives particularly concern the last criterion listed above. The potential launch systems are:

1. The European **Ariane 5**, which can inject at least four satellites directly into their final orbital positions.
2. The Russian **Soyuz/Fregat** for one or two satellites.
3. The Russian **Zenit**.
4. The Russian **Proton**.

Besides rapid implementation of the initial constellation, it is important to have a cost-efficient system for single satellite replacements. Another important issue of satellite design is the process of starting from the overall required reliability of the entire satellite to perform as specified to derive the reliability contributions from subsystems and components and respect them in optimized strategies for the implementation of appropriate system redundancies. This is usually performed in quantitative analyses, albeit taking into consideration that the necessary large number of events which are mandatory to determine the underlying statistical effects are not sufficiently available. As a result, the high power amplification chain for the E1 signal, for example, comprises three solid-state power amplifiers (SSPAs) of which only two are active while the third is engaged only when one of them fails. The satellite system block in Figure 8.6.5 below shows several other examples for redundancies. Further redundancies are used on the equipment and circuit levels as appropriate.

8.6.3 Platform and Subsystems

8.6.3.1 Structure and Thermal Management

The carrier of all components i.e. the physical structure has to mitigate elegantly the contradicting requirements of light weight (low mass), high form stability, and high thermal and electrical conductivity of all satellite components. As already mentioned, **thermal conductivity** is essential for transporting heat to the radiating walls, while the interface to the launcher has to carry all mechanical loads (vibration, shock, acceleration) during the very demanding launch phase. In addition the structure must allow easy access to all critical components, test points, and propulsion load and drain points until the final go-ahead for the

countdown. The structure of the Galileo satellites consists of a metal-laminated cuboid honeycomb structure with **dimensions** $2.50 \times 1.20 \times 1.10\,m^3$. At launch with a full tank of 100 liters, the satellite`s wet mass is about 700 kg at maximum. Integrated supporting elements at critical structure positions absorb excessive loads as can occur in the solar array drive mechanisms and the launcher interface. Metal stringers (titanium–aluminum–vanadium) with very low thermal conductivity ensure a thermally isolated area for the clocks, which include those meeting rubidium atomic frequency standards (RAFS) and superstable passive hydrogen masers (PHMs). Each of the walls serves a special purpose. For instance, the components of the **propulsion system** are mounted on the *x*-wall, while the inner wall carries four momentum wheels which are symmetrically mounted around the center of gravity, as shown in Figure 8.6.3. By means of the **momentum wheels**, the on-board computer can turn the satellite into any desired attitude. The $-z$-side, later facing permanently toward outer space as well as toward the launcher dispenser interface, as shown in Figure 8.6.4, forms the bottom plate of the satellite comprising three segments that offer access to the compartments housing platform avionics, the propulsion tank and most of the internal payload equipment.

Physical extensions of the radiation area of both $\pm y$-panel surfaces of about 15 cm ensure a thermal balance, especially for the high dissipation heat contributed by the high-power SSPAs. The $+z$-wall, eventually the Earth-pointing face of the satellite, carries the entire antenna farm and, on the inner side of the panel, all equipment with low heat dissipation. Depending on the heat generation, some of the equipment is mounted directly onto the wall, other equipment on thermally isolated stringers. Other measures to support appropriate thermal management are painting the equipment's surface black or keeping the open metallic surface depending on whether it acts as a heat sink or source of heat dissipation. Thermally representative satellite models are investigated by computer simulations (CATIA®) to determine the thermal performance of the equipment under all operating conditions; for example, when exposed to solar heat conditions, or during cool-down during eclipses. Payload and platform modules are designed as stand-alone units to enable separate pre-integration and functional verification, with both modules mated at a later stage in the production process. Besides providing the advantage of easy access to both modules during integration, this means that they can also be produced at different sites by teams with different expertise. The central and especially reinforced **load-bearing structure** of the satellite also provides on its *z*-side the separation interface to the launcher or launcher dispenser structure which has to carry heavy structural stress as well. The Galileo satellites must comply with single and multiple satellite launch interface conditions. There are **six isolated thermal zones** in a satellite, each with its own independent thermal design specification and each separately controlled and monitored:

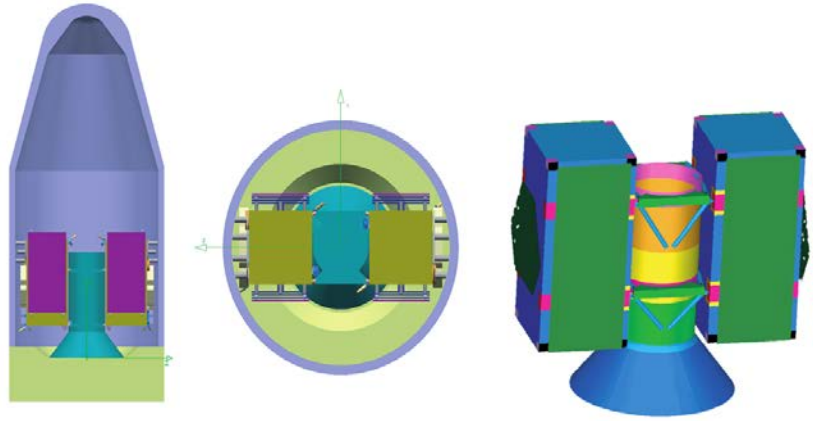


Figure 8.6.4: *Launch configuration with dispenser for two satellites, Soyuz 2 and 1b (Source: ESNIS).*

- Zone 1: clock generator section comprising two rubidium clocks and two passive hydrogen masers.
- Zone 2: all other payload equipment.
- Zone 3: all platform equipment excluding batteries and some components of the propulsion subsystem.
- Zone 4: batteries.
- Zone 5: thermally critical propulsion elements.
- Zone 6: antennas and surface-mounted sensors.

All equipment undergoes comprehensive qualification tests to demonstrate its functional performance under all environmental conditions representative of the operational conditions during launch and on-station operations. Particular attention is devoted to zone 1, which hosts the most sensitive PHMs and is separated by a tent of multilayer insulation (MLI) blankets. Each clock is mounted on thermally isolated stringers on the $+x$-wall, which is never exposed to direct sunlight during its operational life.

Due to its orbital altitude of 23 620 km, each satellite is significantly exposed to space radiation (**Van Allen belt**) created by highly energetic electron densities from solar flares and galactic cosmic radiation. The radiation can cause latchup effects in electronic circuits and memories with inadequate protection. The radiation intensities vary with the 11-year cycle of solar activity. Protective measures include utilization of specifically space-radiation-hardened devices as well as additional protection through shielding walls of high-specific-density material (e.g., tantalum). The latter method can only be applied to components with high radiation sensitivity because of the overall mass constraints. The early Galileo satellites carry radiation dosimeters to determine the actual radiation conditions in the selected orbits over longer periods to avoid overspecification of follow-on satellites.

8.6.3.2 Primary Power Supply

Two solar power generators each produce 850 W (end-of-life) DC power to supply all platform and payload equipment. Batteries with capacities of about 30 A h consisting of cells manufactured using lithium-ion technology ensure uninterrupted power supply to the satellite subsystems during solar eclipses as well as during the early launch and injection phases. During the long ascent phase caused by the required direct in-orbit injection, the closed launcher fairing blocks the solar illumination. The batteries provide the primary power through an electronic power conditioner that maintains a constant DC **bus voltage** of 50 V under all operational load and environmental conditions. As an exception, three devices (the hydrazine valves, the Sun sensors and the magnetic torque device) operate with 28 V. Critical elements of the power subsystem are the two solar array drive mechanisms (SADMs) conducting the solar current into the batteries and ensuring fine tracking of the Sun while the satellite body turns around the y-axis during each orbit. As mentioned earlier, the power subsystem has particularly to maintain battery power for the clocks and other sensitive devices in the satellite during the **long direct injection ascent phase** when the solar array is stowed to fit within the surrounding fairing. A minimum charge current will, however, be available after jettison of the fairing and when the last launcher stage begins to rotate around its velocity axis. During this phase none of the high-power-consuming devices operates except the heaters, in order to maintain the minimum specified thermal conditions for critical equipment. The **DC power distribution** on-board is implemented by a star-shaped cable tree. Resetting electronic fuses protect the fixed power lines from short-circuit effects while all switchable power lines are protected by not automatically regenerating fuses. During all ground operations DC power is provided to the satellite through an externally accessible power connector. Particularly careful attention is required for the DC supply that fires the explosive devices (squibs) which release a holding clamp band when the launcher stage has reached the final orbital position, so that the satellite can be released.

8.6.3.3 On-board Computer and Data Bus

The on-board computer, actually the integrated control and data unit (ICDU), controls the entire status of the satellite by means of sensors and actuators and communicates it via RF telemetry links to the Galileo control centers (GCC) in Oberpfaffenhofen (Germany) and Fucino (Italy). The ICDU is the **central unit for controlling and monitoring** all

platform housekeeping and payload functions. For this purpose, the ICDU interfaces directly with the telecommand receiver and the corresponding telemetry transmitter and controls the other equipment and functional elements through an **MIL 1553B data bus**, as shown in Figure 8.6.5 below. The CPU is an ERC 32 bit processor providing sufficient power with growth capabilities for real-time applications and extended satellite autonomy. An input/output (I/O) module provides access to all sensors and platform units, valves, etc. The ICDU communicates also through the 1553 bus with special security units dedicated to the platform and the payload to ensure high-security protection against malicious access to the satellite. In addition to the 1553 bus, a **space-wire data network** enables switching of redundancy units in the platform and in the payload equipment. One important function of the ICDU is to determine the instantaneous attitude of the spacecraft from inputs of the Sun and Earth sensors. If a change of attitude is commanded from the ground the ICDU software will control the appropriate actuators and momentum wheels to achieve the new attitude.

8.6.3.4 Telemetry and Telecommand

The telemetry and telecommand system (TM/TC) connects the globally distributed TM/TC stations to the two dedicated GCCs in Oberpfaffenhofen near Munich and Fucino near Rome and to the satellites via RF communication links in the 2 GHz frequency band (S-band). Through these links the two centers control the housekeeping functions of all Galileo satellites in space. Telecommand and telemetry signals from TM/TC stations are handled by two helical receiver and transmit antennas located on opposite corners on-board the satellite and the TM/TC transponder (see Figures 8.6.2 and 8.6.5). To receive the telecommand signal the transponder demodulates the BPSK-modulated telecommand signal of 1 kbit/s NRZ-L and channel encoded and source encoded with a standard BCH block code. The transponder receiver delivers the data stream to the ICDU via the platform security unit. The telemetry stream information is delivered at a bit rate of 20 kbit/s NRZ-L encoded to the same antennas for transmission to the TM/TC ground stations. In accordance with the international CCSDS standard, the data content is protected with a **Reed–Solomon (RS) convolution code** to detect and correct single bit errors. All encoding applied to the TM/TC links is to protect them against bit errors resulting from noise and interference on the RF propagation channel. Both links for TM and TC can be operated either in the **ESA standard mode** or in the relatively new direct sequence spread-spectrum mode (also standardized by ESA). By radio command, the TC uplink and the TM downlink can be interconnected in the spacecraft to provide a direct data return link. In this configuration the two interconnected channels provide an active means for determining the slant range from the ground station to the connected satellite. This mode enables determination of the satellite's position and thus, when repeated at later dates or from different ground stations, allows derivation of the Kepler data set necessary to predict the orbit. For Galileo this methodology is not sufficiently accurate. Therefore, in the Galileo system, a sophisticated global ground mission and sensor network, including laser ranging stations, determines the high-precision orbit parameters. Before entering the ICDU the telecommand data is decrypted in a classified hardware environment to avoid unauthorized intrusion into the command links. Two receive and transmit antennas with an **omnidirectional antenna characteristic** (−3 dBi) are mounted on both the lower and upper side end positions of the satellite to ensure reliable access of the TM/TC ground stations to the satellite under all attitude conditions. This is particularly important in emergencies. A comprehensive software package enables a variety of different transmission modes for the required telecommand, such as single commands, file transfer and other specific data structures.

8.6.3.5 Attitude and Orbit Control

All satellites of the constellation need to be kept in their nominal **orbital positions** to maintain an equal distance among them in the orbital plane as this ensures an equal distribution of all satellites around the globe. For the same reason, the satellites in the neighboring orbital planes must remain in their positions to maintain the optimum phasing in the constellation. Because the implementation concept of Galileo relies on direct injection into the final orbital position, the

satellites do not need a powerful propulsion system for orbital maneuvers. With the small 10 N thrusters and 70 liters of monopropellant fuel (hydrazine), a satellite creates a total Δv of about 180 m/s over its lifetime, which is sufficient to drift eventually into its final graveyard orbit. The three-axis attitude control ensures permanent orientation of the $+z$-axis toward the center of the Earth (**nadir pointing**) while the solar arrays remain **Sun pointing** during each orbital revolution. Two Sun sensors, one each for coarse and fine tracking of the Sun, and two Earth sensors provide all the information required to determine the actual attitude. Three redundant computer-controlled wheels with large rotating masses orient the satellite around the three degrees of freedom (axes) and are able to achieve any required attitude position. A comprehensive software package for attitude control takes over immediately after separation from the launcher. This unique software package provides all control and monitoring functionality and can be activated by ground control commands. One of the first maneuvers after separation and housekeeping tests is to deploy the solar arrays to ensure sufficient availability of primary power. The software also provides autonomous activities, for example, in case of an emergency, by changing the attitude into Sun acquisition mode as the **fail-safe mode** and alerting ground control to take appropriate action. Besides keeping the satellite in Sun-pointing mode to maintain electrical power supply, in this case it would also be important to maintain stable thermal conditions in each area and to configure the other systems of the satellite in a standby mode to support satellite-saving measures. A globally distributed network of 10 ground stations is located such that at any moment in time each satellite is in direct radio line of sight of at least one station.

8.6.3.6 Laser Ranging Reflector

For ultra-precise ranging and orbit determination, the satellites are equipped with a **passive laser ranging reflector** (LRR) located on the Earth-pointing surface ($+z$). Because the overall achievable ranging accuracy should be in the order of less than 1 m over a slant range distance of about 26 000 000 m from a ground terminal to the satellite, the instantaneous position of the satellite varies slightly due to the slightly varying gravity conditions at each point. Laser ranging enables determination of the instantaneous slant range to the satellite in the order of a few centimeters, albeit only during optical visibility conditions. More information about the global network of several laser ranging stations is provided in [8.6.6].

8.6.3.7 Propulsion

Due to the very limited number of satellite orbital maneuvers, the propulsion system comprises a **monopropellant** (hydrazine) system with a spherical tank of about 100 liters volume and four fully redundantly provisioned 1 N thrusters. The tank is mounted on an enforced frame structure to cope with the high dynamic forces during launch and separation. Pipes made of special titanium alloys deliver the fuel from the tank to the thrusters through electrically controllable valves. For safety reasons, during the launch campaign a "safe"/"arm" connector accessible from outside the satellite prevents the unintentional firing of thrusters. This connector eventually is set to the "arm" position as one of the final steps in the countdown procedure after the spacecraft is fully loaded with fuel, mounted in the fairing and ready for launch. The propulsion system also has a drain outlet to evacuate the tank during qualification tests when the tank is filled with nontoxic replacement liquids, and also in case of an aborted launch.

8.6.3.8 Wire and Cable Harness

Today, the wire and cable harness still forms the central nervous system of a satellite. There is a growing trend to use serial data buses for the data and control communications, as in industrial and home computer networks. On-board the Galileo satellites the serial 1553B bus is applied, which is not optimized for satellites but which has evolved over many years into a more or less *de facto* standard. Other buses, such as the CAN bus widely used in the automobile industry, is used in one of the two preoperational Galileo satellites (GIOVE A). However, the great leap in mass saving with an optical or wireless LAN on-board has not yet occurred, mainly due to high RF power-level conditions in the vicinity of the spacecraft – but who knows what will happen in the future? The heaviest wires are

those which carry the primary power to high-power consumers such as the HPA RF stages. Another option is to use the metal surface of the satellite for the ground return, if a metal surface is actually provided to cover the honeycomb structure.

8.6.4 The Galileo Payload

8.6.4.1 Payload Architecture

The payload architecture as shown in Figure 8.6.5 provides the following performance features:

- Generation and conditioning of the high-precision navigation **clock signals** created on-board.
- Receiving, demodulating, decoding and storing uplinked **system data** such as ephemerides, almanacs, timing and time-offset values for clock correction in users' receivers and others.
- Receiving and sequencing **integrity data** uplinked by the integrity determination ground networks.
- Composing different **navigation messages** for the navigation services (open service, SoL, public regulated service and the commercial service) including forward error correction decoding.
- Generating the different **spreading codes** for the signals.
- Particular **encryption** of the coding for the security signal (public regulated service).
- Conversion of the baseband signals to the actual **RF transmission frequencies** as well as power amplification for transmission by the navigation antenna with isoflux characteristics.

To provide these functions the payload comprises the following equipment:

- **Mission receiver** with its own dedicated antenna in the 5000–5010 MHz band (C-band) for the reception of mission-relevant data for navigation messages and control functions of the payload.
- **Ultrastable oscillators** (atomic frequency standards) in rubidium and, very ambitiously, passive hydrogen maser technologies (RAFS and PHM).
- **Control and monitoring unit** (CMCU) for the oscillators and their redundant units to process the 10.23 MHz and the 10.00 MHz high-precision frequency clocks from both units.
- **Decryption unit** to secure authorized access to the payload.
- Navigation signal generator unit to create mathematically correct **signals in the time domain** for the multiplexed E1, E5 and E6 signals.
- RF carrier oscillators **synchronized** with the mother clock signal.
- Wideband 50 W and 70 W RF **linear power amplifier** for the L-band frequency ranges of 1164–1300 MHz and 1559–1594 MHz, respectively.
- High-power **output multiplexer** for common transmission of the E5 and E6 signals as well as the direct through-path for the E1 signal.
- The main **navigation antenna** transmitting all signals, which requires extremely wideband characteristics combined with isoflux characteristics. This is to ensure less RF power in the boresight direction pointing toward the subsatellite point and more RF power dedicated to the edges of the coverage.
- **Remote terminal units** to provide the connecting interface to the computer for many sensors, actuators and switching devices.
- A stand-alone transponder with its own receive and transmit antenna system listening on the **international distress calling frequencies** in the 406 MHz band, as well as a transmission antenna for the radio link to the COSPAS-SARSAT ground network (local user terminal).

8.6.4.2 Payload Equipment

C-band Mission Receiver

The mission receiver receives the spread-spectrum modulated and coded signals transmitted by a global network of mission uplink stations in the **C-band, 5000–5010 MHz**. It is thus ensured that seamless, robust and reliable access to the payload is possible at any moment in time. The receive antenna is implemented as a right hand circularly polarized (RHCP) horn antenna with a comparatively low gain to provide a wide angle of capture for uplink stations. After the received signal is downconverted, demodulated and despread, it is routed through a security unit into the navigation signal generator where its content is processed and submitted to the navigation message encoder and sequencer.

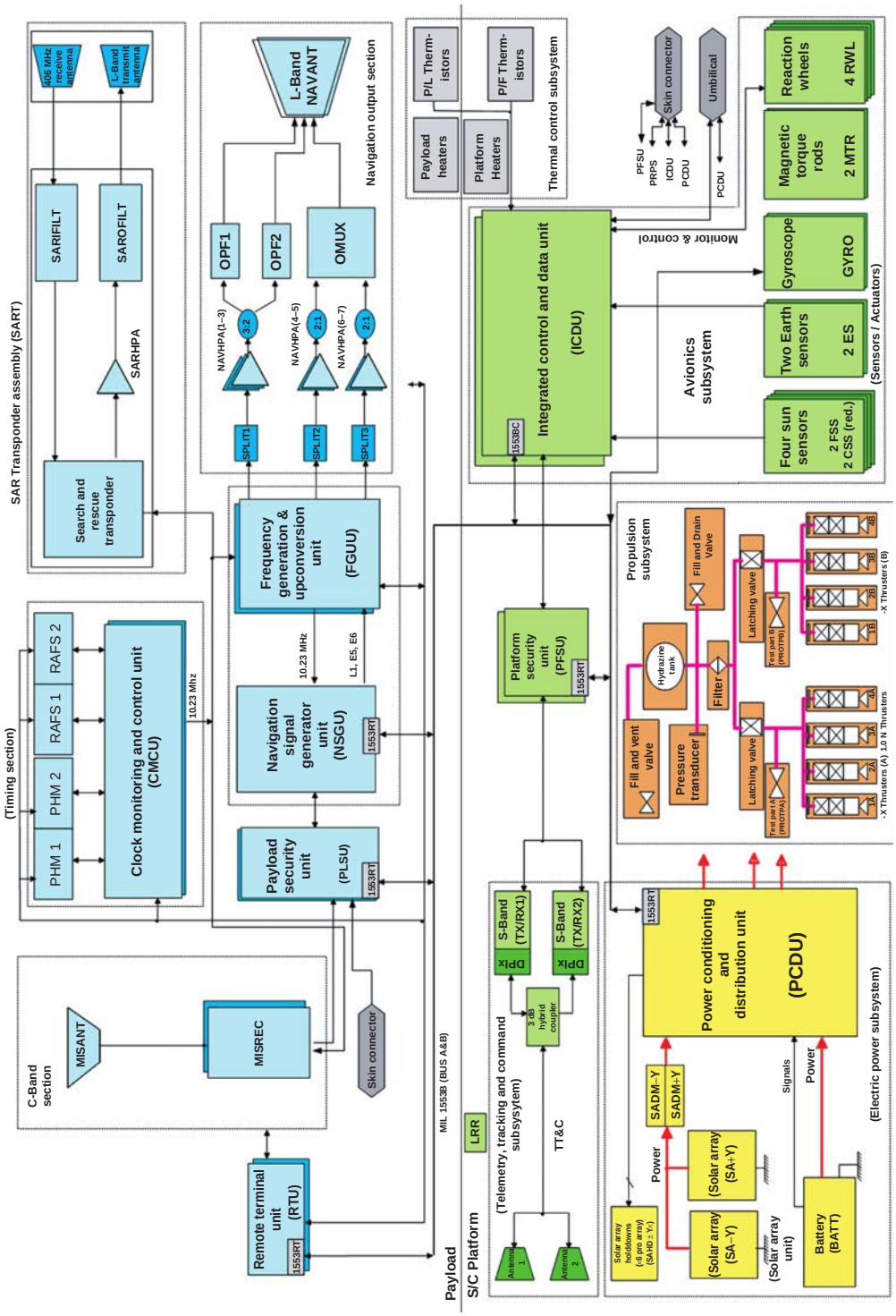


Figure 8.6.5: *Architecture of payload and platform equipment (Source: ESNIS).*

Ultrastable Frequency Standard Generators
The core of all navigation satellites is the on-board **ultrahigh-precision clock generator**, selected from a group with different technologies and redundant standards. They are also called "clocks," although they actually provide a precise frequency signal or pulse trains and not "real time-clock" information. A real time clock with hours and minutes is derived from such a clock generator by a time definition of the Galileo on-board and system time. Real time is internationally defined by convention of time and date. This time is defined in the ground station and synchronized to the international reference of coordinated universal time (UTC) by counting the seconds to higher units (minutes, hours, etc.). The ground system time is compared to the actually disseminated signal time from space and the instantaneous time error (ns) is transmitted via the navigation message. With this information, receivers correct the instantaneous time error by "software" rather than by adjusting the hardware. Galileo satellites apply two different technologies for the clocks: the more traditional **rubidium atomic frequency standard** (RAFS), used on-board GPS satellites for many years, and two highly sophisticated **passive hydrogen masers** (PHMs) developed by European industry under ESA contract specifically for this application. The clock generator generates a time signal to the highest precision regarding phase noise, frequency stability and lowest drifts. The frequency stability of RAFS is in the order of 10^{-13} s while the PHM can achieve a long-term stability of 10^{-15} s over 24 h. Any of the four clocks on-board the satellites can create the master clock signal, which is selected by the clock monitoring control unit (CMCU) and used to derive coherently all RF carriers, subcarriers and mixer frequencies for upconversion and data rates.

All clocks run continuously, providing the master clock as selected by ground control. **On-board clock accuracy** is controlled on the ground with an ensemble of high-precision clocks at the locations of many standard time service providers. Deviations are not corrected by physical clock corrections but through parameters of a clock model taking all impacts including the relativistic compensation into account. After physical changes it takes several days to restabilize the clock for use. By applying reverse modeling the navigation receiver can calculate the real instantaneous arrival time of the phase coherent signal to a high precision [8.6.7]. The on-board master clock is conditioned and distributed to a variety of on-board payload equipment and to the linear converter inside the SAR transponder. In total there are actually five independent clock-synchronized systems on-board; all of them are eventually synchronized with the Galileo system time. This ensures that any reasonable combination of Galileo signals in the connected receivers (e.g., single- and dual-frequency receivers) will lead to valid position solutions.

Navigation Signal Generator
The navigation signal generator unit (NSGU) generates three composite signals named "E1," "E5," and "E6" which are permanently transmitted from all Galileo satellites. Each signal comprises two (E5) or three (E6, E1) components that alone or in combination can be used by a wide range of receivers for low- and high-end applications. Data message content uplinked from the ground control centers via the connected uplink station is decomposed in the NSGU and recombined (added) at bit rates from 50 to 250 bit/s to the pseudo-random noise (PRN) stream. The PRN-coded subcarriers applied for the Galileo signals are spread by direct sequence spreading up to 15 million chips. The NSGU generates all the signals at baseband level, as shown in Figure 8.6.1. By means of different binary offset coding (BOC) schemes the signals have different spectral power distributions, which is desirable for sharing frequency bands among the many signals of the same constellation as well as among signals of different constellations. This is very important since not only do other RNSS systems share the frequency bands, but so too do other radio services such as primary radar. With the extremely wide bandwidth spreading (e.g., 50 bit/s over 32 MHz) Galileo actually produces nearly white-noise-like signals which cause **minimal interference** with other systems, if any.

Carrier Generator and Frequency Upconverter
All baseband signals delivered by the NSGU are upconverted to the three final dedicated RF carrier frequencies in the following frequency generation and upconversion unit (FGUU). A local oscillator synchronized to the high-precision standard signal (10.23 MHz) generates the carrier signals, followed

by a mixture stage that converts them to the final frequencies. All generated signals applied to the up-converters to create each of the three carrier signals are intermediate-frequency (IF) signals which themselves are also synchronized to the master clock signal. Each of the navigation signals, E1, E5 and E6, is multiplexed with the associated navigation message transmitted at effective symbol rates ranging from 50 to 1000 bit/s. Since the redundancy bits ensuring the required **forward error correction** (FEC) capability require an overhead of 50% of the symbol rate, it leaves the other 50% for net data information. This capacity enables transmission of all mandatory system information, comprising almanacs, ephemerides, time error and geodetic reference data, as well as integrity information and acknowledgment messages for distress calls and more, differing for each signal.

RF Power Amplifier

The readily composed signals on the exact RF transmission frequency are linearly amplified to the specified output power level through an ensemble of redundantly configured SSPAs. The E5 and E6 signals in the lower frequency band of 1164–1300 MHz with effective RF bandwidths of 51 and 40 MHz, respectively, are routed through 70 W RF amplifiers configured in a 2-for-1 redundancy. The amplified signals are then applied to the antenna through an output multiplexer. The E1 signal with an RF bandwidth of about 35 MHz is amplified to an RF output power of 50 W through a 3-for-2 redundancy SSPA configuration. This means that the output power is provided by two parallel SSPAs selected out of three. Optimized bandpass filters in each branch of the power amplification stages reject out-of-band emissions and harmonics to avoid harmful interference with radio services in neighboring and harmonic frequency bands. The **radio astronomy service** is particularly sensitive to interference and thus uses very large high-gain antennas. The primary power provided by the solar arrays is designed to ensure permanent operation of all three signals.

Main Navigation Antenna

The main transmit antenna for the navigation signals consists of an array of radiators, which by appropriate phasing of all feed elements creates an RHCP wavefront and the required isopower flux characteristics. The isoflux characteristic means that the received power level at a received reference antenna located on the Earth's surface stays nearly equal if the antenna is located at the center of the radiation (boresight equal to subsatellite point) or at the edge of the coverage. This special characteristic is achieved for both frequency bands, 1164–1300 MHz and 1559–1594 MHz, by two physically implemented **beam-forming networks** mounted below the feed array. The antenna design is quite sophisticated as some of the elements contribute to both bands. The antenna gain is adjusted to the coverage requirements of that antenna which serves the entire visible surface of the Earth. The **antenna characteristic** has lower gain at the center (corresponding to 90° elevation from a ground antenna) and higher gain at the edge of coverage (defined by 5° elevation from a ground antenna). The system is defined and designed so that each satellite provides a signal receive power of typically −122 dBm at a "0 dBi" antenna located anywhere on the surface of the Earth, at any time. It is very important for service quality to provide exact phase relations among all transmitted signals, which is ensured as explained above. In addition, the satellite design ensures that the phase center of the antenna is less than 20 cm from the center of gravity of the satellite. This minimizes the impact on the dilution of precision caused by the rotation of the satellite around the *z*-axis during each orbit, as explained in Section 8.6.3.5.

8.6.4.3 Search and Rescue Payload

The SAR transponder on-board the Galileo satellites will provide a major improvement regarding timely availability of the international air and sea rescue services offered under the intergovernmental COSPAS-SARSAT agreement among France, the USA, Canada and the Russian Federation, with more than 30 associated ground segment providers, user states and international organizations such as ICAO, IMO, ITU, and others. Galileo supports the services of this organization to "assist search and rescue activities on a worldwide basis by providing accurate, timely, and reliable alert and location data to the international community on a nondiscriminatory basis" [8.6.8] by providing full and permanent seamless global

coverage for distress calls. Full global coverage is provided by only half of the Galileo in-orbit fleet for SAR access. The Galileo SAR transponders receive distress signals from registered distress beacons (EPIRB) on the internationally allocated frequency band of 406.0–406.1 MHz and retransmit the linearly converted uplink signals in the band of 1544.05–1544.15 MHz that is also protected by international conventions for SAR purposes (ITU-R). So-called **local user terminals** (COSPAS-SARSAT LUTs) receive the distress calls that are repeated through Galileo (and other GSO and non-GSO) satellites and generate a dedicated acknowledgment message to the particular distress-calling terminal (**emergency position-indicating radio beacon** (EPIRB)) through the navigation message in the Galileo open service signal. By notifying the caller that the call is affirmed and that appropriate rescue actions are underway, including brief messages about the measures taken, the caller is made aware of the status and can stop any further transmissions, leading also to more efficient use of the sensitive distress call channels.

The **SAR antenna** is an array of six quadrifilar helical receive antennas mounted on an electrically conducting ground plane with a transmitting backfire L-band antenna in the center for the downlink, as shown in Figure 8.6.2. The feeds of both antennas are phased to achieve RHCP for the receive and LHCP for the transmit antenna. The uplink receive RF bandwidth of 90 kHz can be repeated in one slice or split into two subbands of about 50 kHz each to be filtered and retransmitted on the downlink. Besides receiving distress calls from a new generation of EPIRBs with built-in GNSS receivers, the system also still supports the use of the **legacy mode** where EPIRBs are located through triangulation by LUTs through signals from several satellites. The legacy mode is still supported by Galileo because many EPIRBs are still in use. However, because of many false alerts it has been decided by COSPAS-SARSAT [8.6.8] that, as from February 1, 2009, only the 406 MHz band will remain for SAR while the other distress frequencies of 121.5 MHz and 243 MHz used so far will no longer be supported. It may also be considered that only EPIRBs with built-in GNSS receivers will be supported, first, as stated earlier, to free the frequency channel for further potential distress calls and, second, and probably equally important, to save battery energy in case more calls are necessary before an acknowledgment is received.

8.6.5 Launcher Interfaces

To rapidly build up the space segment of the Galileo system, during the deployment phase the majority of satellites are to be carried into space by **multisatellite launches** using Ariane 5. To achieve this, a decision to place more (10) satellites in fewer (three) orbital planes as compared to GPS with its six orbital planes was taken in 1999 after completion of an ESA comparative systems study. To complete the constellation and for later replacement of individual satellites, single launches with other systems such as Soyuz with the Fregat upper stage, Zenit or Proton rockets are also possible, given an appropriate satellite design. For multisatellite launches, **dispenser** structures support the carriage and safe separation, enabling the launchers to place each satellite into its final orbital positions, as shown in Figure 8.6.4. The dynamic and environmental conditions required to comply with several different launcher interfaces have a large impact on the design of the satellites. Exacting tests to confirm the survival of the satellites in the fierce environment of the launches are mandatory. This is discussed in the next section.

8.6.6 Satellite Assembly, Integration and Testing (AIT)

8.6.6.1 Introduction

The following sections discuss the stages of development and manufacturing of the main components, equipment and subsystems, and the process of verification of all specified parameters as well as the qualification for achieving flight approval. The satellites comprise two separately integrated and tested modules. These are the **platform module** with all the subsystems required to provide and maintain the mandatory electrical and environmental operations conditions and the **payload module** carrying all the flight components of the navigation system which is the purpose of the mission. Having two main modules

enables separate integration and test activities, which leads to significant time and cost savings. Due to the relatively high number of satellites to be produced for the Galileo constellation, the satellites and their modules must also be designed to comply with requirements for ease of manufacture. Therefore, two integration sites, one for payloads and one for platforms, will tentatively be contracted by ESA acting on behalf of the EU for **series production**. The potential sites selected so far are organized in consecutive production "cells" where assigned mechanical and electrical integration steps are performed. Accordingly, each cell is equipped with appropriate tools, work skills and test equipment. Each satellite, starting with the carrying structure, travels on a movable dolly from one cell to the next for completion, thus allowing the specialized teams, their tools and the test equipment to remain at their cell positions.

8.6.6.2 Model Philosophy

Nearly all payload equipment and some of the platform systems have been developed especially for the Galileo navigation mission, taking into account the space radiation conditions for space vehicles travelling at an altitude of nominally 23 616 km. The development cycle undergoes similar model steps for all the space equipment. A distinction is made between models for functional tests, models for software development, models for interface verification and, last but not least, models for environmental tests to achieve qualification for space flight. After integration of the components into the platform or payload modules, tests only on the module level will be performed since the individual pieces of equipment will already have proven their qualification for space flight. In the frame of the four-year cycle of system and equipment development, the so-called "in-orbit validation (IOV)" phase, each piece of equipment developed for Galileo passes through four model stages, STM–EM–PFM–FM:

- The **structure and thermal model** (STM) to verify the mechanical interfaces for fixing points, cable harness routings, mass and volume under specified environmental conditions.
- The **engineering model** (EM) to develop and verify all electrical interfaces (functional hardware and software interfaces).
- The **prototype flight model** (PFM) to undergo all the tests for approval of space qualification (temperature, vacuum, electromagnetic compatibility (EMC)), vibration, acoustic noise, etc. The PFM comprises all redundant parts and may even be used in some cases as the later flight model.
- The **flight model** (FM) to be assembled in a qualified manufacturing process and applied to all satellites in the constellation.

Depending on the complexity or the technology of a piece of equipment, several models of the same category may be necessary. This is the case for instance for the newly developed RAFS. Besides the development of on-board equipment, further variants of the same equipment are developed, for example for transportable satellite simulators, so-called suitcase equipment. These simulators are mainly used to test and verify the performance and the various RF interfaces with the ground stations. The suitcase equipment receives and transmits 2 GHz TM/TC signals with representative protocols and "receives" mission uplink messages with correct protocols in the 5 GHz band. Additional RF interfaces allow the SAR transponder functions to be tested. Similar to the component model philosophy used by the various suppliers, different satellite models support the development of the satellites before series production of the flight models commences. The details of these models are as follows.

Structure and Thermal Model (STM)
The STM allows **design decisions** to be verified, such as the accommodation of equipment, placing of the fuel tank, valves and pipes, sensors, cable harness routings and fixings. Even though the entire design of the spacecraft evolves with the help of CAD programs on a workstation, a real mass model remains important to obtain information on mechanical, dynamic and thermal loads at each corner of the satellite. The STM also serves to verify the launcher interface and the separation system.

Engineering Model (EM)
The EM is functionally a complete satellite, albeit not approved for flight. Therefore, the EM usually **verifies all interfaces** with the ground segment and all the performance parameters required by the spacecraft

specification. The interfaces to be tested and verified comprise the RF links to the mission uplink and TM/TC stations as well as the direct access wire link (**umbilical**) to the on-board computer. The umbilical cable enables direct access and satellite control independent of RF link access. Beyond this, the EM provides representative structural and equipment conditions for investigating and verifying the protection offered by the installed **RF shielding** and the measurement of bonding conditions on all metal surfaces. All metal surfaces must be interconnected with the lowest resistance **bonding** possible to ensure equipotential among all conductive elements of the satellite and to neutralize potentially high voltages that can build up when the satellite travels through areas of charged particles emitted by solar flares. Low-resistance bonding prevents the build-up of differently charged surfaces in one part of the satellite which could result in sudden high-voltage discharges. These discharges could potentially destroy equipment or the entire satellite. Good bonding and shielding are also important prerequisites for preventing different electrical potentials and high-voltage peaks in the vicinity of high-power devices (SSPA). Furthermore, good bonding is essential for avoiding ground loops which can cause humming and sometimes unstable operating conditions in the equipment and subsystems. Last but not least, the EM is used to test and verify the important interface with the **electrical ground support equipment** (EGSE) to be certified for use with flight hardware.

Prototype Flight Model (PFM)

The PFM is a complete satellite built according to **full flight standards** to test the effect of all the environmental interface conditions that are expected during launch and in orbit, including the highest expected levels. A PFM comprises all the components, redundancies and materials of a flight model. Although not fully flight qualified, since it has been comprehensively tested the PFM might be launched as an additional flight unit however with reduced lifetime expectations.

Flight Model (FM)

The FM comprises only **flight-standard components** and redundancies and is subjected to lower test profiles confirming more the manufacture of the assembly rather than verification of the design. In case of launch delays, Galileo FMs are stored under controlled environmental (clean-room) and power conditions. Although FMs are stored without batteries, external power supplies ensure **permanent stand-by operation of the high-precision clocks** (PHM). Even during the transport of the satellites, clean-room and operations conditions must seamlessly continue until the satellite is mounted atop the launcher. The clean-room conditions even continue until the fairing is jettisoned into free space when contamination by particles or liquids can no longer occur. In the Galileo program with its 30 operational FMs in space, each satellite will undergo a reduced test program because all processes for the final assembly and integration line have been previously validated and proven to be repeatable, ensuring consistent manufacturing and system quality. At the time of writing, the actual conditions for the series production of the Galileo satellites are still being determined because the European Commission, as the system owner and operator, has not yet made all decisions concerning the additional deliveries needed to complete the system. The presently ongoing development phase (the IOV phase) will eventually result in **four satellites in space** with all the essential key ground system elements operationally necessary to prove the overall system concept. In the context of the IOV phase, further test environments will be developed and verified for use in the full deployment phase of the program. These are as follows.

A Special Measuring Station to Determine Payload Performance

This test site involves all key payload components mounted on a partially complete carrying structure to verify the key payload performance parameters. This comprises all interface signals and connections to the platform module as well as to all payload equipment. Dedicated **special checkout equipment** (SCOE) allows each payload unit to be appropriately operated. The test results achieved at the station already form part of the overall data package to achieve flight qualification and will thus not be repeated in the course of the final satellite integration. A Qualification Review Board (QRB) including representatives of the customer, the prime contractor (and relevant subcontractors) and the launch service provider confirms flight

readiness in a declaration of design and performance compliance.

A Software Validation Facility (SVF)
Although most of the algorithms for attitude control and for remote control (TM/TC), as well as the drivers for sensors and actuators and for all the other standard satellite functions, actually exist, a large number of adaptations and changes must be implemented and tailored for this mission. The SVF provides all the facilities (computer, interfaces, tools, etc.) to test and verify any additions and modifications to the flight software. Together with the satellite simulator, the SVF is a permanent instrument that follows the spacecraft throughout its lifetime and provides a controlled environment for any change or upgrade of the **flight software** in accordance with the applicable quality procedures. Any software change is first verified on this facility before it is uploaded to the satellites in space. This both ensures that only valid software gets uploaded to the satellite and guarantees that approved and configuration-controlled software is loaded on the on-board computers.

Avionics Test Bed (ATB)
Following the practice of the aviation industry to consider all on-board equipment as "avionics," the satellite industry also tends to call the **attitude and orbit control system** (AOCS) "avionics." The complex test environment needed for the development and closed-loop verification of algorithms and subsystem performance is modeled with a special avionics test bed facility. An ATB enables the simulation and testing of algorithms for attitude control and their actual implementation by stimulating the Sun and Earth sensors as well as the actuators and wheels to apply physical forces and momentum. Both normal and nonnominal operational conditions can be simulated in order to investigate algorithms, software and corresponding peripherals during nominal and emergency conditions. An ATB comprises representative components of the Galileo AOCS subsystem: a nonflight but electrically representative flight computer (ICDU); all coarse and fine sensors, wheels and gyroscopes; and the simulators representing the Sun and Earth as "seen" from any real position in the orbital planes. The ATB is completed by a suspension harness allowing full 3D movements to stimulate and investigate AOCS performance under representative dynamic conditions.

8.6.6.3 Integration and Platform Tests

The model and modular test philosophy breaks up the design and development phase of the satellites into several activities performed at separate places with dedicated test systems and skills. The overall project management coordinates all activities with regard to the schedule which in turn is driven by the customer and the launch date for a satellite or group of satellites in the case of multisatellite launches. The development program for the satellites foresees the following steps toward final release for flight (launch):

- **Mechanical and structural approval** of the satellite design (in a preliminary design review followed later by a critical design review, PDR/CDR)
- Proof of **thermal balance** in all anticipated nominal and emergency operational situations of the satellites
- Proof of the **electromagnetic compatibility** (EMC) of the satellite and its subsystems, again under all operational conditions
- Detailed verification of all **performance parameters** as required by the satellite specification [8.6.3] under all environmental conditions (for platform and payload)
- Verification of the **flight software** and all interfaces with the satellite subsystems and ICDU periphery
- Verification of all **satellite interfaces** with the corresponding ground stations (TM/TC, mission uplink stations, sensor stations, etc.).

The complete **verification program** is only applied to the first four satellites (IOV phase). It is not repeated for each follow-up satellite unless necessitated by a major change to the satellite. The flight qualification is applied to all satellites that are built to the same manufacturing specifications for the equipment and the satellites. The emphasis of testing during the manufacture of series satellites concentrates on the skills and processes to ensure high quality under continued flight clearance. However, samples are taken and submitted for closer reassessment regarding qualification status. In the case of major changes affecting

the qualification status, a requalification process (**delta qualification**) is performed. Any changes and their justifications must be seen in light of the impact on cost and schedule. All the described measures not only serve to verify specifications and avoid risks, but also form the basis for the justification of possible insurance claims if a failure occurs during the in-orbit lifetime of a satellite. An overview of the typical tests performed on each of the models described is provided in Table 8.6.1.

8.6.6.4 Payload Tests

Similar to the platform test programs, equivalent payload tests are performed as integration proceeds toward completion in order to demonstrate the specified payload performance under all anticipated environmental conditions. However, some differences need to be explained. As shown in Figure 8.6.5, the payload contains the ensemble of atomic clocks controlled by the CMCU, followed by a frequency carrier generator and the navigation **signal generator** which creates the transmission signals (sequencing, encoding, modulation, multiplexing) on **baseband** frequencies. The **complex signals**, comprising the required 10 signal components as shown in Figure 8.6.1, are linearly and coherently upconverted in mixer stages by using appropriately derived harmonics from the master clock signal. The frequency upconversion stage delivers its composite signals, on final transmit frequencies, through the SSPAs to the navigation antenna. All stages have to demonstrate compliance with their specifications before the entire navigation payload is verified. The **search and rescue payload**, although otherwise independent of the payload, has an important interface with the clock ensemble by receiving a coherent and high-precision conversion signal. All equipment undergoes detailed performance verification tests under the suppliers' control and responsibility, also because only the suppliers have the appropriate skills and special test equipment. The main objective of the payload tests is to verify that all equipment is working correctly **end to end** and produces navigation signals that comply with the payload specifications under all space environmental and electrical conditions. The results are used to decide whether the individual components and their specifications sufficiently fulfill the actual requirements.

In planning the test campaign, it is important to take account of the high RF radiation power, the behavior of the antenna characteristics under test conditions, the sensitivity of the clock generators to variations in the ambient temperature, and several other aspects to achieve flight approval for the payload. Some of the equipment is checked by automated routine measurements to determine behavior under long-term operating conditions. One practical example of long-term observation is the monitoring of the current drawn by each individual component during nominal operations and during state changes. The recurrence of a typical current profile is a good indicator during the production phase of equipment and payload behavior to quickly verify that adequate quality (product assurance) has been obtained. Trends, the occurrence of unexpectedly high variations or short glitches in current consumption can provide good indications of a critical condition within a unit. An integrated data collection system (IDAS) enables all performance data taken during the development and lifetime of equipment and payload to be compiled and analyzed. Examples of data collected include:

- The **bonding conditions** to verify the continuing low-resistance interconnectivity of electrical ground planes and metal surfaces.
- The actual **electrical mating condition** of multi-pin cable connectors.
- The **power supply** (total power consumption, voltage stability, inrush currents, power consumption of individual pieces of equipment).
- The signal quality of commands and other signals regarding voltage level, bit rates, pulse shapes, frequency offsets, etc.
- Voltage levels of the acquired **analog sensor** information.
- Verification of acquired values with specified/expected values and corresponding tolerances and **calibration** curves.

Dedicated test systems (EGSE) enable the device to be tested in any desired mode of operation (by stimulation or simulation) to test its performance in the detail required. Modalities and comprehensiveness of the test programs and test load conditions

Table 8.6.1: *Satellite models and corresponding tests.*

Test	STM	SM	EM	PFM	FMs
Incoming inspection	X	X	X	X	X
Electrical, functional and performance test:					
• integration and test			X	X	X
• integrated system test (IST)			X	X	X
Quasi-static load test with sine-shape vibrations, low frequencies in three axes	X				
Acoustic test (launcher-dependent noise profile)	X			X	X
Vibrations test (sine shape) for resonances	X	X		X	X
Verification of all mechanical interfaces: shape, volume, mass, fixings, cable and pipe fittings (fit check)	X	X		X	X[a]
Separation shock: simulating shock during separation from launcher or dispenser	X			X	
Microvibration test to determine vibration impact during on-station operations	X			X[b]	
Deployment test of all solar array doors				X	X
Solar illumination test (before and after environmental tests)				X	X
Temperature cycling in static vacuum	X			X	
Performance under cyclic variation of thermal vacuum conditions				X	X
Simulated tanking and draining with fuel substitute	X			X	
Electromagnetic compatibility (EMC) tests: conducted emissions on cable harness (mainly on power but also on signal cables)			X	X	X[c]
EMC tests on conducted susceptibility on signal and data lines as well as receiver inputs				X	
Impact of electrostatic discharge (ESD) on cable harness and connected equipment			X		
Impact of electrostatic discharge (ESD) on receiver input stages			X		
Verification of the unified propulsion system (UPS) and subsystems				X	X
Verification of alignments and calibration of critical values (where applicable)	X			X	X
Determination of effective dimensions (important for transport and launcher fairing interfaces)	X	X		X	X
Tests to determine interference and coupling conditions among all transmit and receive antennas of the antenna farm on the +*z*-panel surface			X		
Tests of RF compatibility between all transmitters and receivers, including determination of passive intermodulation products (PIMs). (A PIM is created by applying high RF power to an environment of joints of different metals, faking diodes that create intermodulation products that can potentially fall within a frequency band used by on-board SAR receivers)				X	
Antenna characteristic after environmental tests				X	X
Tests to validate the entire satellite system (SVT)			X	X	X

[a] Only with models of equipment built for fit checks (represented in form, mass, volume, attachments, connectors).
[b] Only emissions on signal cables, not the potential impacts in case nominal values are exceeded.
[c] Full payload "end-to-end" test with all equipment turned on.

comply with the conditions that the equipment will experience during its launch and in-orbit lifetime. Actual operating conditions vary significantly with the actual location of the equipment within the satellite. The most stressing and lifetime-relevant payload tests are concentrated on the EM, while the tests on FMs are limited to those that do not impact the required in-orbit lifetime of at least 12 years. The main objective of the limited production tests is to detect potential initial component failures which could occur in the early operational phases of equipment. Military equipment even undergoes dedicated burn-in tests to separate delivery equipment from the equipment requiring retrofit before being used in the field. The test program for the FM payloads comprises:

- Measurement of the **clock accuracy** regarding signal level, clock stability (Allan deviation), phase noise behavior, autonomy.
- Determination of the actual characteristics of the **on-board antennas**, namely navigation antenna (phased array) (1.1–1.59 GHz), mission antenna (5 GHz), SAR antennas (406 MHz and 1.544 GHz) through dedicated test points.
- Measurement of **RF characteristics** of the transmission signals regarding RF output power (min/max level and stability), frequency accuracy, signal bandwidth of all RF signals, degree of unwanted out-of-band emissions, phase relations of transmitted navigation signals and the achieved wavefront (phase center). Further, phase noise and passive intermodulation (PIM) under full-power transmission conditions caused by "diode" joints of different metals in close vicinity to the antenna farm are detected.
- **Optical alignment** of the antenna mounted atop the payload module. The final high-precision alignment is performed on the satellite level.
- Determination of the required isoflux **radiation characteristic** of the mounted antenna. This is performed in a special near-field antenna test facility under normal operational conditions because this is important for the overall ranging accuracy achievable by this payload (satellite). Tests with follow-up antennas use the test signals derived from the generic measurements at the facility to verify the performance.
- Testing **resistance**, particularly to conducted and radiated emissions (**EMC**) from the high-power on-board transmitter. These measurements are performed in special test rooms (**Faraday cages**) with high shielding protection against transmissions from outside.
- Verification tests of the **SAR payload** regarding the system receiver noise performance, the bandpass characteristics of each converted channel (bandwidth, ripple and group delay performance), RF output power, signal return delay time, and power consumption.

8.6.6.5 Integrated Spacecraft Tests (IST)

After successful completion of all subsystem tests on the payload and platform levels, and the integration of both modules, the integrated spacecraft system tests follow with the solar array and batteries mounted. On completion of the tests with the fully integrated system it is ensured that the entire satellite complies with the specifications and that it will successfully survive the launch and in-orbit (radiation) conditions for at least 12 years.

The final system tests are the functional verification tests for the entire satellite, controlled and monitored by its own internal TM/TC facilities rather than through the umbilical line connection. The results of these tests are compared to the results of earlier tests. The clock generator tests are restricted to nominal functionalities because it always takes a long start-up time to bring them to stable operating condition. During the system tests the primary power is supplied by the actual solar arrays (stimulated by solar simulators illuminating each sector of the array) and the batteries. The system test comprises final EMC tests concentrating on those cables and wires that are installed only in the integrated satellite.

The complete satellite is finally connected through remote control (TM/TC) data communication lines (ISDN/VPN) to the GCCs for the so-called **system validation test** (SVT) to ensure full system command capability. For this test, all satellite functions are turned on in full operational configuration (including activation of all RF power transmissions). The ISDN (or virtual private network (VPN)) connects the GCCs to the EGSE, which in turn communicates

through its RF interfaces with the satellite under test. Finally, the completed satellites (PFM and FMs) must prove their operational stability under varying **thermal vacuum** conditions. This test verifies the thermal balance of each thermal zone in the satellite under varying ambient "space" conditions imposed by different simulated **solar illumination and eclipse conditions**. It is important to verify by means of the thermal control system that no point within the satellite or in any one of its components will experience overheating or excessive cooling. The thermal balance conditions verify the absence of excessive mechanical tensions which otherwise would be caused by thermal gradients exceeding the specified limits. The active thermal control system is based on a **heat pipe architecture** that conducts excessive heat particularly from the high-power amplifiers to the heat radiation walls of the satellite. The ultimate objective of the ISTs is to verify that the designed and manufactured satellites fully comply with all specified requirements and are therefore ready to serve their purpose.

8.6.6.6 Launcher Integration and In-Orbit Testing (IOT) and Validation (IOV)

With the transport to the launch site, which tentatively occurs soon after completion of the satellites, the last phase before the actual beginning of the in-orbit operational phase commences. It is important that, during the entire transport and storage period from the final integration site to the launch pad, clean-room conditions are maintained inside the acclimatized transport and storage containers, even during integration of the satellite into the launcher fairing. To verify that the satellite successfully survived the entire transport, some key performance tests including battery conditions are repeated before integration of the satellite with the launcher.

Monitoring of the basic performance parameters is continued while the satellite is waiting for launch. In case of irregularities in these key satellite parameters, the countdown can be suspended until a few minutes (6 minutes for Ariane) before liftoff at any time, as soon as the satellite mission director declares the satellite to be in a "no-go" condition. One of the characteristics of navigation satellites is that they can only demonstrate their full 3D positioning fleet performance when at least four satellites are visible from a location on Earth. Compliance with specifications and the correct functioning of each satellite are the necessary preconditions for the Galileo system. This is confirmed as soon as the joint ranging performance of all satellites contributes constructively to a position and navigation solution with an accuracy that complies with the mission specification [8.6.1]. This last test is the key test for the **proof of concept**, thus validating the correctness of the entire system concept. When this last hurdle is surmounted, validating the correct interoperation of ground and space segments, the deployment of all the remaining satellites will commence, after which the declaration of **full operational capability** (FOC) of the Galileo system will establish the beginning of regular services.

Bibliography

[8.6.1] Galileo Mission Requirements Document (GMRD), Issue 6.0. European Commission (EC)/European Space Agency (ESA), July 26, 2004.

[8.6.2] Galileo Signal Interface Control Document (ICD) describing the free-to-air navigation services. European GNSS Supervisory Authority. http://www.gsa.europa.eu/.

[8.6.3] Galileo System Requirement Document (GSRD). ESA Version 4, rev. 2, July 24, 2004.

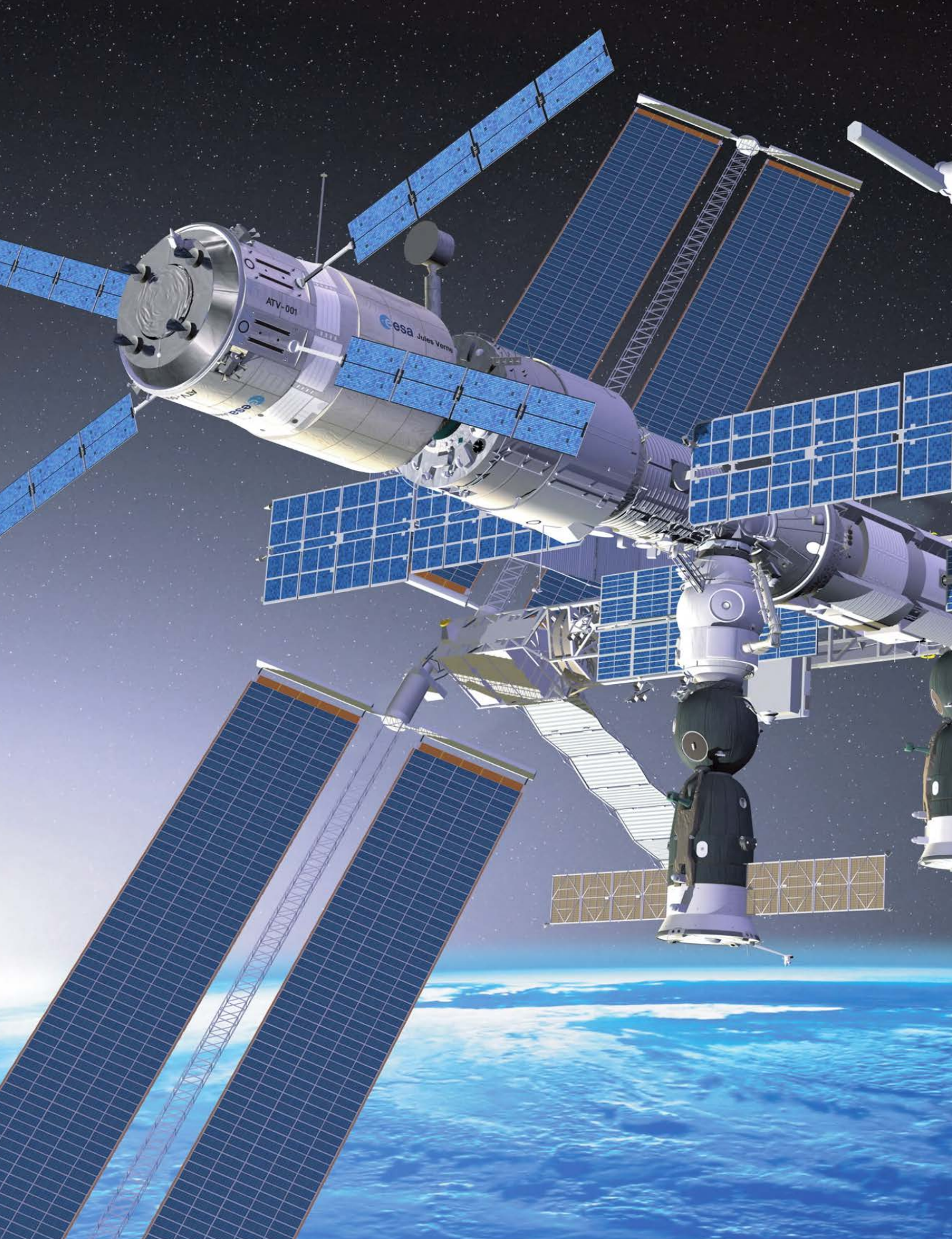
[8.6.4] European Cooperation for Space Standardization (ECSS). http://ecss.nl/.

[8.6.5] Consultative Committee for Space Data Systems (CCSDS). Concept Papers. http://public.ccsds.org/publications/ConceptPapers.aspx, 2008.

[8.6.6] More details of the satellite laser ranging network of ground stations can be found in http://ilrs.gsfc.nasa.gov/stations/index.html, 2007.

[8.6.7] Misra, P., Enge, P. *Global Positioning System – Signals, Measurements, and Performance*. Lincoln, MA: Ganga-Jamuna Press, 2001.

[8.6.8] More information about the SAR service actually supported by Galileo can be found at http://www.cospas-sarsat.org/, 2008.


ATV-001
esa
Jules Verne

9 Management of Space Projects

Wilfried Ley

The management of space projects has to guarantee the successful execution of a project, so that the following goals are achieved [9.0.1]:

- Adequate engineering
- Acceptable and transparent costs
- Timely accomplishment of scheduled milestones.

In this context a project will have the following characteristics [9.0.2]:

- A unique and therefore noncyclic order of events
- A defined starting and termination time
- Distinct task descriptions and objectives
- Participation by teams, working groups, companies and institutions
- High complexity.

The processes allowing the project manager to achieve the required goals for the essential elements of a project during its total life cycle are described in the following sections:

9.1 Management of space projects
9.2 Quality management
9.3 Cost management
9.4 Legal aspects of space activities.

In each case appropriate risk management needs to be applied.

The essential project elements are:

- Project structures: Technology coordination, administration, documentation and finances; competences and responsibilities; planning, time calculation, cost calculation and supervision
- Project organization: Preparation of project management plans
- Project phase and planning: Development of milestone plans for controlling the project's progress
- Configuration management: Realization, description and control of the system's technical aspects during the life cycle
- Information/documentation management: Provision of all necessary information
- Cost management: Optimum utilization of staff at all the institutions involved and also of materials and finances
- Schedule management: Preparation of milestone plans and their continuous supervision.

Any failures are classified according to predetermined criteria, in order to have a qualitative basis of assessment in a form that might permit future use [9.0.2].

Class 1 Catastrophic:

- Failures which cost human life or cause severe injury or are otherwise extremely detrimental to health.
- Failures which cause the loss of a spacecraft.

Class 2 Critical:

- Failures which cause injury to or health problems for participants without seriously endangering them directly, but requiring medical treatment.
- Failures which cause a substantial reduction in the mission goals.

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9

Class 3 Considerable:
- Failures which cause minor injuries or minor health problems without negatively affecting the availability of the participants.
- Failures which cause only minor reductions in the mission goals.

Class 4 Negligible:
- All other failures.

Possible failures with catastrophic and critical consequences (classes 1 and 2) are significant and must be corrected immediately (e.g., by redundancies).

Bibliography

[9.0.1] ECSS-M-00A. *European Cooperation for Space Standardization, Raumfahrt-Projektmanagement.* DLR, Qualitäts- und Produktsicherung, Nationales ECSS-Sekretariat.

[9.0.2] Hallmann, W., Ley, W. *Handbuch der Raumfahrttechnik*, 2. Auflage. Munich: Carl Hanser Verlag, 1997.

9.1 Management of Space Projects

Joachim Klein

An overview of project management tasks and methods used for the realization of space programs is provided here.

Based on typical space technologies, methodologies and applications, the main features which must be taken into consideration for successfully implementing a space project or program will be presented. While the previous chapters of this handbook described the basic feasibility, the characteristics of the technologies and the underlying physics, this section deals with the methods of project management implementation and the efficient control of tasks executed by the different parties involved in a space program.

To fulfill the aims and objectives of this handbook only a generic overview of project management will be given, supported by typical examples.

9.1.1 Project Management

9.1.1.1 History

Without going into a detailed consideration of prehistoric and ancient times, a few well-known historical ventures of interest in the context of this presentation should be acknowledged. To name only a few, the legendary Tower of Babel, the Pyramids of Gizeh, the Great Wall of China and Gothic cathedrals should be recalled. All these ventures had the following in common:
- A strong and untiring political will
- A challenging and complex task
- The long duration of their realization – decades or even centuries.

Even with today's knowledge and criteria, the efforts invested by these former architects and builders in coordinating and mastering the processes involved are more than remarkable. They were unique and even today are partially unknown.

The development of space technologies in the 1950s and 1960s had similar attributes. These can be illustrated by the Apollo program:
- Political goals in the Cold War between the USA and the former Soviet Union
- The complexity and difficulty of a human mission to the Moon
- Interdisciplinary and intercultural cooperation.

And in contrast to former enterprises there were:
- Stringent and timely requirements
- Strict requirements for safety and the protection of human life and health
- Limited or controllable risks during realization.

From the experience gained in several realized programs, including their successes and failures, and using the Apollo missions as an example, it has become evident that such missions can only be carried out successfully with a consistent, well-directed and coordinated management approach. Since that time, nothing has been left to chance.

Quality *is never an* accident*: It is always the result of intelligent effort.*

John Ruskin (1819–1900), British philosopher

In the past few decades all the tasks and processes required for the realization of a space program have been systematically developed. The present space industry is able to make use of an increasing pool of experience, documented in various international and industrial standards.

While the success of former space missions was primarily oriented along technology lines and performance values, nowadays success is more and more measured by profit considerations and value for money. The definition of project tasks is changing from the management and execution of **development programs** to the realization of economical enterprises, in particular the **production of assets** based on known and qualified processes and methods.

This makes it evident that the expectations of a space mission are no longer limited to the fulfillment of technical and scientific requirements. Due to the tight and limited budgets of public authorities, consumers, agencies and institutions, the profitability of a program and its management approach toward accomplishing the project goals within the requested time and cost frame are also of essential importance.

Success or failure of a program will to a lesser extent be determined by its technical challenges than by its management accomplishments. Many failed or collapsed programs may be due rather to mismanagement than to failure to meet technical challenges.

This conflict can be illustrated by the Cluster satellites, which should have been launched into orbit after their successful manufacture and qualification with the first launch of the Ariane 5 rocket. Though the launch failed for technical reasons, public opinion judged the failure to be caused by the wrong **management decision** to launch the four satellites using a newly developed and not finally qualified launcher.

The reverse conclusion is that successful project management determines the success of a program. What are the **success factors** for successful project management and what are the tasks?

These questions will be discussed and illustrated below. Although they focus on space applications, they are also applicable to other projects and programs in different technical domains with different quality ratings and contents.

9.1.1.2 Introduction

Considering the above and in retrospect, projects in the space business may be seen as enterprises typically characterized by the following aspects:

- Uniqueness of the implementation approach
- Time limitations
- Limited resources
- Political goals
- Risky processes
- Intercultural and multicultural cooperation
- Interdisciplinary challenges
- Highly complex requirements and tasks.

Project management which is tasked with the execution of a space program is faced with the tension that exists between all the project partners, as illustrated in Figure 9.1.1.

Project management (PM) interfaces with all the partners, in particular with:

- The customer (CU), who controls the contractual requirements of the procurement.
- The executive management (CEO) of the procurement organization.
- The project team (PT), which is tasked by executive management with the manufacture and delivery of the product to the customer or end user.

Project management has to deal with the various interests of the project partners and to do justice to all of them in order to achieve the mission goals of the procurement contract.

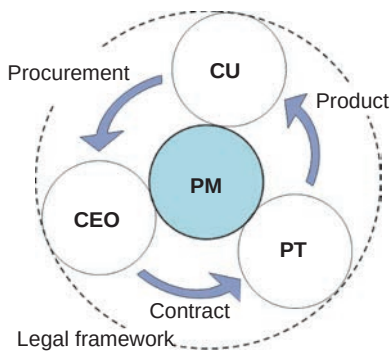


Figure 9.1.1: *The context of project management's role.*

Table 9.1.1: *Tasks and interests of the project partners.*

	Project tasks Contractual frame	Interests
CU	• Definition of requirements • Provision of sufficient financial resources • Contract partner • Project control and status monitoring • Confirmation of project progress • Release of interim results • Review of documents and commenting by RIDs • Payments • Acceptance of deliverables	• High level of product quality • Smooth project execution • Minimization of effort
CEO	• Contract partner • Provision of resources (personnel, material) • Tasking of the project team • Authorization of the project management	• Management of project risks • Profitable project execution without financial loss
PT	• Execution of the project tasks • Respecting project requirements	• Assurance of employment • Satisfaction during execution
PM	• Management of project • Delivery of product	• Harmonious project execution

At first glance this trivial relationship includes many individual tasks and interests of the various project partners, all of whom have to be coordinated to achieve the mission goals within the contractual framework.

The interfaces of project management with the project partners define the qualification profile of the project manager. Basically the profile is divided into three parts and includes capabilities and competences in the following areas:

- Functional, technical, product-oriented (**technical competence**)
- Management know-how, methodology, process knowledge (**methodological competence**)
- Interaction with people, leadership, communication behavior (**social competence**).

Only a sound balance between the areas of competence vis-à-vis the project partners and their interests will guarantee a successful project.

What are these areas of interest and what are the **tasks of the project partners**? With respect to these questions Table 9.1.1 provides a generic overview of typical project management tasks.

It is noteworthy that all the tasks of the project partners are on the one hand specified and defined within the contract, and on the other hand should seamlessly complement each other. The interests of the parties, however, may conflict and accordingly influence and disturb homogeneous project execution. For example:

- The desire of the contractor to achieve a product of the best quality with a minimum investment of resources with respect to the budget and available time.
- The desire of the contractor to maximize profit and return on investment during project execution.
- The desire of the project team to maintain an even distribution of workload during project execution.

Most of these are certainly incomplete and imprecise; however, they illustrate what must be considered by project management for a successful project execution.

While the **social competence** of the project manager is important for balancing the individual interests of the project partners, the manager's **technical and methodological competence** is essential for the successful execution of the project for the benefit of all partners.

The quality of the deliverable product as defined by the requirements of the procurement contract

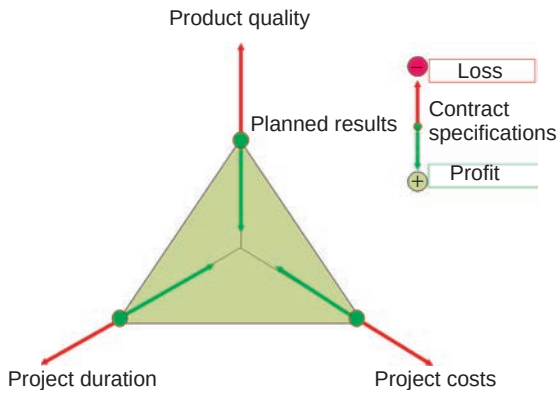


Figure 9.1.2: *Triangle of success for project management.*

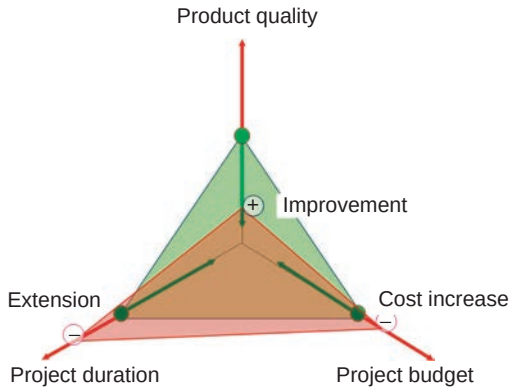


Figure 9.1.3: *Effects of an improvement in product quality.*

directly depends on the project cost and schedule (see Figure 9.1.2).

Assuring an acceptable balance between the **success factors** of product quality, schedule and cost is the core task of the project manager. The figure also shows that a change in one of the parameters will affect the other parameters and may threaten the overall success of the project compared to the original planning.

Figure 9.1.3 shows the effect on costs and schedule of an assumed quality improvement of the product. As a consequence, interference in the relations between the involved project partners can be assumed as well, as indicated in Figure 9.1.3.

The causes of this interference are manifold and are interdependent. However, their effects on the success of the project are mostly negative.

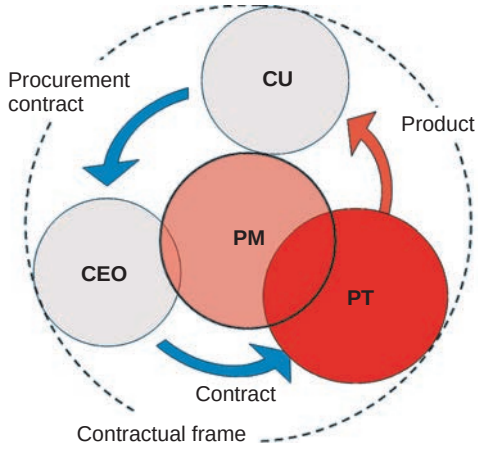


Figure 9.1.4: *Conflicts of interest between project partners.*

In this sense, project management is also understood as the totality of leadership tasks for successful execution of projects. The responsibility associated with the project management task explains not only its attraction, but also its challenges (Figure 9.1.4).

In reality, projects are commonly characterized by complex structures and multilayered task descriptions, combined with the risk that bad subtask solutions or minor changes may cause dramatic follow-up problems.

Project management must therefore focus on a system-oriented approach and take into account systems-level technical relationships: "Thinking in systems, the holistic view, is an effective approach inherent in project management". It is supported by **system engineering management** as a complementary subdiscipline to avoid the related project risks.

In large projects, a major part of the traditional tasks of project management may be transferred to the system engineer, who integrates all the technical engineering disciplines and capabilities into a consistent, team-orientated, structured process and efficiently supports and complements the work of the project manager.

In Figure 9.1.5 the competence profile of a project manager versus the size or scope of a project is shown in a qualitative manner. The figure shows which capabilities and competence a project manager should have to successfully lead and guide project tasks,

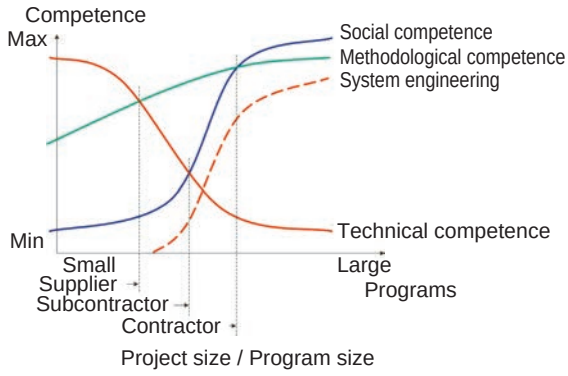


Figure 9.1.5: *Competence profile of project management.*

acting as supplier, subcontractor or contractor of a space-related program or project.

While for small tasks (e.g., supplying components) project management should have high technical competence, for large projects this competence may be less important because of the integration of system engineering functions in the project architecture as an independent task.

Despite what has been said above, the requirements for social competence and the associated tasks increase with increasing size and complexity of a program.

The whole is more than the sum of its parts.
Aristotle (384–322BC)

This guideline formulates the main goals for project management:

- The manufacture of complex products requires the cooperation of different organizations with a common target: delivery of a space product consistent with customer requirements within a given time and cost frame.
- Technical tasks, personal capacities and financial resources must be structured, organized and managed (coordinated).
- Project management comprises the definition, execution and implementation of tasks in compliance with the requirements and needs. This includes the verification process.
- Project management requires detailed investigations and analyses of what is to be realized and in which way. This includes planning the necessary detailed steps and the resources needed.
- Most important, however, is coping with reality and with the problems, delays, changes, hurdles and alternatives which may arise during the project.

Table 9.1.2: *Fields of knowledge of project management.*

Competence	Examples
Methodological competence	• Risk management • Procurement management • Schedule/time management • Cost management • Integration management
Technical competence	• Quality management • Content interface management
Social competence	• Human resource management • Communication management

These targets are achievable in the most efficient way if the activities of project management with respect to the project partners are dominated by the following characteristics:

- Transparency and directness
- Commitment to delivery and agreements.

Project management in general covers specific fields of tasks and expertise [9.1.1], which are allocated to several competence fields of a project manager (see Table 9.1.2).

The fields of expertise listed in the table are within the job qualifications of a project manager and are described in detail in the next sections to the extent that they are necessary for and characteristic of a space project. The focus will be on the fields of knowledge of methodological competence. The fields of knowledge assigned to technical competence are specific to a project and are described in the various chapters of this handbook. The fields of social competence are only touched upon.

9.1.2 Space Project Characteristics

The success factors of a project, such as schedule, cost and quality, depend in general on the requested tasks, the size and the complexity. Project durations of

several weeks to many years are possible, with related project effort and costs.

The success of a space project or any supporting part depends essentially on the detailed definition of requirements and the resulting target definitions. Only then can the execution and implementation steps between the beginning and end of a project become visible and predictable. This detailed planning is the baseline for the control tasks of project management.

This becomes critical if, as is usual in space programs, several project parties, organizations or companies are involved and must be guided and controlled.

Reality shows, however, that the necessary planning effort is often minimized or even neglected. Excuses are typically: "The management responsible already knows what to do" in the case of smaller projects; or for bigger projects, "Planning uncertainties are still too high because of expected changes."

Therefore, the efforts of management services must be efficiently tailored for economic reasons. This leads on the one hand to different expectations of the competence of project management (see Figure 9.1.5) and on the other hand to formal consideration of the content and scope of all tasks (see Section 9.1.4).

This conflict between the necessity for detailed planning and the tailoring of needs can only be solved by applying systematic and sequential procedural methods which relate to the specific characteristics of the space project or the deliverable space product.

It is strongly recommended that "working-level" project view is taken and that the so-called **top-down approach** is used for the planning and related changes, taking into consideration the **life cycle** of the product.

For better understanding, a complete project cycle will be discussed below, starting with the initial development of a concept and ending with fulfillment of the project goal or the use of the product.

The nature of a project may differ, being either a **development** or a pure **procurement** program. While procurement programs make use of commercially available and qualified goods and processes with adequate heritage of demonstrating their functionality in space missions, development programs may require additional process steps and effort.

Commercially based programs are normally associated with the procurement of communications satellites, ground stations and appropriate operational capabilities. Development programs are normally associated with the development and manufacturing of satellites with new technologies or new functional performance.

In particular the **model philosophy**, determining among other things the number of models to be manufactured as well as the contents of verification measures, may be seen as a compromise between effort/costs and acceptable program risks.

Commercial procurements commonly accept known risks for the benefit of smaller investment.

The correlation (in principle) between effort and risk is shown in Figure 9.1.6.

9.1.2.1 Processes in Space Technology

All processes between mission definition and operation or use of a product are subject to a common logic, which may be seen as a system approach. It is schematically presented in Figure 9.1.7.

In this closed loop the customer initiates the project by providing sufficient financial resources and delegates responsibility for the project to the main contractor through contractual regulations and technical procedures. At the end of the value-added

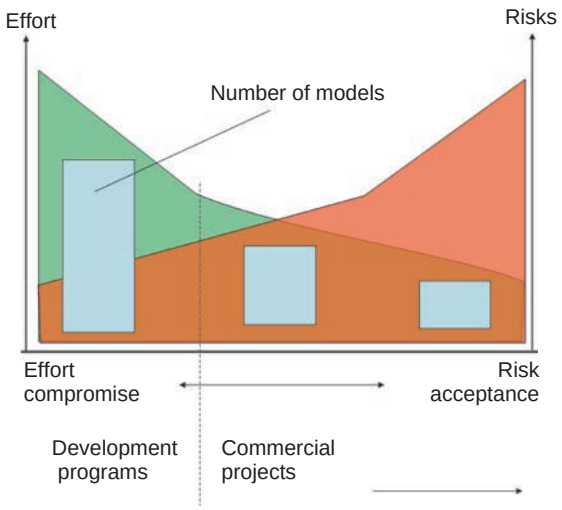


Figure 9.1.6: *Relationship between effort and risk.*

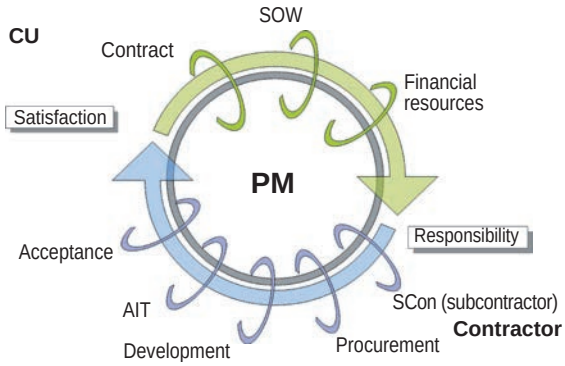


Figure 9.1.7: *Quality-oriented process understanding.*

production chain the realized product can be accepted to the satisfaction of the customer. The roles and responsibilities may be distributed among internal or external partners.

This understanding, based on modern quality management approaches, assures a harmonic balance of interests in successful projects and is essential for the following reasons:

- **The Customer's Shared Responsibility:** The customer has to define the requirements and expectations as clearly as possible and has to provide reasonable and sufficient financial resources. This sounds trivial but, as a matter of fact, it is often the subject of misunderstanding or even legal disputes in cases of wrong or incomplete definitions and requirements.
  Often those deficiencies lead to suspension of a project or even to bankruptcy of a contractual partner. In any case this will have effects on the quality of the product and will influence the satisfaction of the customer or user.
  The customer is therefore advised to offer assistance in all project phases to jointly reach the mission goals, to review and control its status and progress, and to accept the final result or product after positive evaluation. Related payments must be result oriented and initiated in a timely manner.

- **The Responsibility of the Contractor:** The contractor's management has to carefully review the tasks using internal guidelines and must contractually confirm the cost and time frame for the planned implementation. It has to be assured that internal resources are available and that quality standards are applied for executing the project.
  Quite frequently different understandings of what constitutes project realization lead to extensive technical or even private divergences between the involved parties and in consequence to qualitatively weak project performance.

- **The Integrity of all Subprocesses:** The overall program consists of multiple processes which have to be carried out and concluded individually. All processes must complement each other and fulfill the overall project goal. Those responsible for a process and the respective tools must be available without restrictions to support the project manager.
  Intermediate processes must be reviewed and accepted at formal reviews and **milestones**. Departures from this understanding lead in most cases to work overloads as well as to resource and cost problems.

- **The Value-Adding Process:** Value-adding processes must be clearly defined. It must be assured that products or services comply with requirements and are acceptable by the customer at the end of the implementation process for all life cycles.
  Wrong or insufficient understanding often leads to undesirable results with consequences for costs and the schedule. Overdesign caused by the designer's own ideas has to be avoided.

- **The Satisfaction of the Customer:** Customer satisfaction is primarily achieved when tasks are performed in compliance with planning and requirements, and secondarily when the customer's expectations and goals are fulfilled (see Table 9.1.3 below).
  Contractors must carefully and systematically plan the needed processes and execute them transparently. This aspect is of considerable importance because "badly" performing projects can often be traced back to deficient understanding of how customer satisfaction is to be achieved.

- **The Overall Task of Project Management:** Project management must be regarded as a central contact point where monitoring and control of the entire value-adding process take place. It must be

pro-active and motivating, and in case of conflicts, moderating and convincing.
The primary task is to assume complete responsibility for the project, contractual project execution and delivery in full to the customer. This comprises all competence fields of project management which, if not adequately observed, may lead to significant financial problems and loss of image for the companies or people involved. Only comprehensive consideration of so-called closed loop process understanding leads to a classic "win–win" situation for all project partners.

However, for the authorities, institutions and companies involved in many space projects, it is normal for the most important processes and management tasks to be strictly regulated and controlled by dedicated **quality management systems** according to ISO 9001, ISO EN 9100 or ECSS.

These are applied during the individual phases of the **life cycle of a space project**, which are characterized as:

- Conception
- Definition and tasking
- Design and development
- Manufacturing and commissioning
- Acceptance
- Operation
- Disposal.

This is also valid for components and subsystems which are assembled and integrated into the systems.

For economic reasons this system approach has to accept phase overlaps and "workaround" solutions (Figure 9.1.8). As explained later, it is essential that each phase is accepted and finalized by reviews. Delta reviews may be systematically integrated into the project systematics and are subordinated to system reviews.

Remark:
In contrast to the project cycle, the **life cycle of space products** according to [9.1.5] is determined by the time period which starts with acceptance and extends until disposal of the product, and includes:

- Transport to the launch pad
- Launch preparation tests
- Integration on the launcher
- Launch into the defined orbit

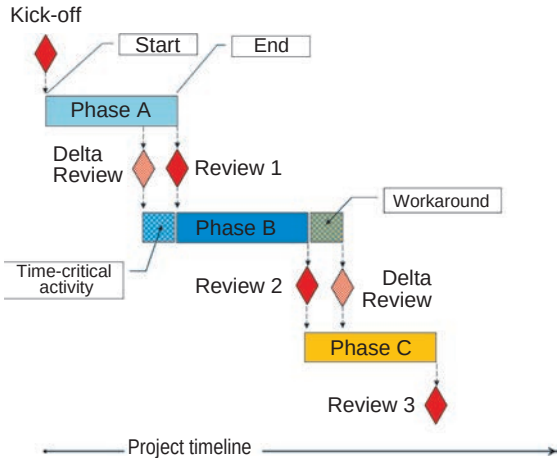


Figure 9.1.8: *Phase overlaps.*

- Operation in orbit
- Disposal by reentry or positioning in a graveyard orbit.

This general representation of a project's processes and implementation phases is realized in the space business by one of two different alternatives which are conceptually similar but differ because of varying overall conditions from one case to another. These are the **ESA implementation phases** and the **commercial procurement approach.**

9.1.2.2 ESA Implementation Phases

The ESA has developed and introduced a system of standards and regulations for planning and implementing space programs and projects which are mandatory for European space programs.

The standards of the **European Cooperation for Space Standardization** (ECSS) are internationally recognized and are comparable with NASA's Mil-Std (military standard) series. They are based on the former ESA PSS Standards and are subdivided into three parts (Figure 9.1.9):

- Engineering standards (E-series)
- Product assurance standards (Q-series)
- Management standards (M-series)

The ECSS series comprises three levels of details:

- **Level 1 Policy and objectives:** These standards define the strategy for specific disciplines, specify

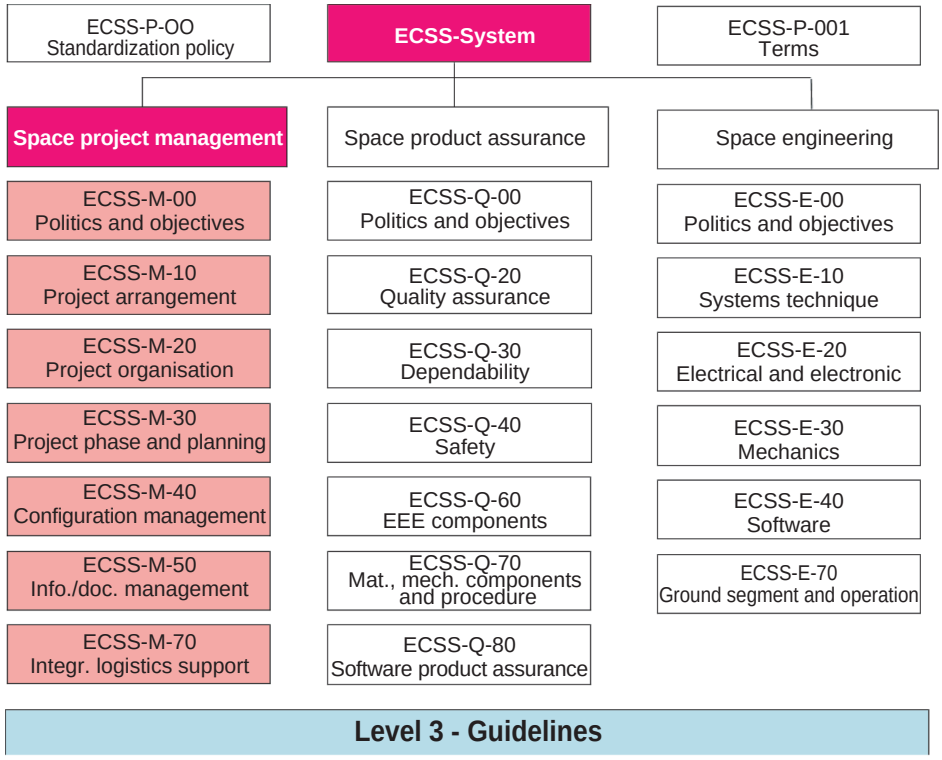


Figure 9.1.9: *Overview of ECSS standards.*

general requirements for specific disciplines, and define the interfaces according to Level 2 standards.

- **Level 2 What is to be accomplished – expected output:** These standards define requirements (goals and functions) for all aspects of specific disciplines (project organization, quality assurance and system engineering).
- **Level 3 How it is to be done (guidelines):** These standards define methods and processes, and provide recommendations for achieving Level 2 requirements.

Remark:
Level 3 standards are not within the scope of this chapter.

The management standards define requirements for project management and describe the procedures to be followed in a space program. The relevant aspects are described in the following:

- **Level 1 Standards**

ECSS-M-00 ***Project Management*****:** This standard describes the basic principles of project management and their integration with product assurance requirements and technical elements for all programmatic phases starting with the request for quotation up to the commissioning of the product.

- **Level 2 Standards**

ECSS-M-10 ***Project Structure*****:** This standard provides guidelines for compilation, use and tailoring of project structures and their implementation within a project.

ECSS-M-20 ***Project Organization*****:** This standard provides guidelines for industrial program organizations, customer structures and internal/external interfaces.

ECSS-M-30 ***Project Phases and Planning*****:** This standard defines basic principles and requirements to be observed and monitored during program execution.

ECSS-M-40 ***Configuration Management*****:** This standard specifies rules for the configuration management

including identification, monitoring, status reporting, definition, verification, and maintenance of hardware, software and documents.

ECSS-M-50 ***Information/Documentation Management*****:** This standard provides requirements for the information and documentation management systems with respect to correctness, completeness, availability, accessibility, security, compliance and status reporting.

ECSS-M-60 ***Cost and Schedule Management*****:** This standard covers the monitoring of costs and schedules.

ECSS-M-70 ***Integrated Logistic Support (ILS)*****:** This standard provides the organizational and management rules of ILS activities for a project, relationships to project management, and the definition of logistic support analysis.`

An overview of the project phases as described in [9.1.6] is given in Figure 9.1.10.

Each phase is concluded by appropriate **reviews** in which a dedicated **review board**:

- Confirms the achieved results.
- Defines follow-up actions.
- Authorizes continuation to the next phase.

Phase 0 Mission Analysis
Goal Analysis of a mission to identify and characterize the planned mission goals.
Review **Mission definition review (MDR)**

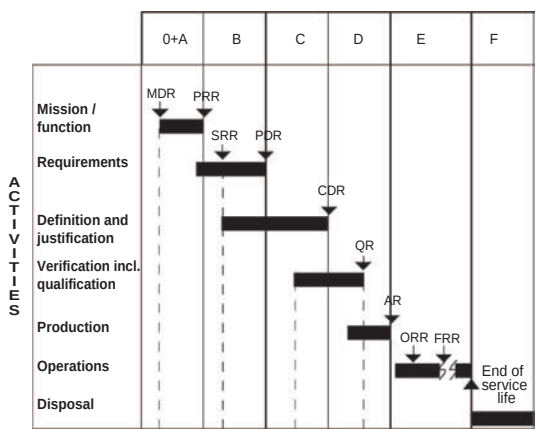


Figure 9.1.10: *Project phases according to ECSS-M-30 [9.1.6].*

Phase A Feasibility
Goal Finalization of the mission characteristics and conceptual compilation of solutions and associated conditions, compilation of different system concepts, finalization of the functional requirements.
Review **Preliminary requirements review (PRR)**

While Phase 0 and Phase A studies are performed by independent or scientific institutions, the following Phase B of the proposed mission is carried out by the main contractor selected by the customer.

Phase B Preliminary Definition Phase
Goal part 1 Preliminary definition of the mission with selection of possible technical solutions for the system concept chosen in the PRR.
Review **System requirements review (SRR)**
Goal part 2 Further detailing, selection and definition of methods, resources and products including estimation of effort and implementation planning.
Review **Preliminary definition review (PDR)**

Based on the interim results of Phase B as confirmed by the SRR, the contractor will be contractually tasked to perform the follow-up phases.

Remark:
With respect to time, the PDR is performed in the design phase (Phase C) (phase overlapping).

Phase C Detailed Definition Phase
Goal Detailed investigation of the selected solution supported by manufacturing and qualification of representative elements, confirmation of feasibility and fulfillment of requirements.
Review **Critical design review (CDR)**

Phase D Production Phase (MAIT Phase – Manufacturing, Assembly, Integration, Testing)
Goal Procurement of components, manufacturing of first models (qualification model) for qualification of the selected design, verification of manufacturing methods and procedures.

Review	**Qualification review (QR)**
Goal	Manufacture of flight models on the basis of qualification test results, verification of reliable manufacturing, proof of functional performance and operation, release for transport to the launch pad.
Review	**Acceptance review (AR)**

Remarks:
The procurement process as well as the component development starts in Phase C. The results of qualification measures on the component level are already reviewed at CDRs.

Phase E	**Operation (LEOP** (Launch and Early Operation Phase) **and Operation Phase)**
Goal part 1	Evidence of functional performance of the overall system (satellite and ground segment), mission rehearsal.
Review	**Operational readiness review (ORR)**
Goal part 2	Preparation and execution of the launch campaign, release for launch.
Review	**Flight readiness review (FRR)**
Goal part 3	Commissioning of satellite and overall system, operation and use.

Phase F	**Disposal Phase**
Goal	Conclusion of the end of life performance, system deactivation.

The project flow shown is also valid for all involved subcontractors, who are subordinated to the main contractor on the system level (**top-down approach**) as shown in Figure 9.1.11.

Carrying out reviews to approve the delivered services remains the responsibility of the customer. The contractor(s) has (have) to provide the requested performance and service inputs. This assures formal approval of the development status and release of the results by the responsible higher level customer.

The customer confirms the project progress achieved and takes legal responsibility for the delivered service. With this positive confirmation the customer authorizes the appropriate payment, which is commonly linked to major reviews.

The ESA approach applies to all space programs of ESA and its institutional customers, in particular for programs with a clear focus on development.

It should be mentioned that the ESA approach may be tailored to project needs related to technical implementation, risk, safety and economic efficiency [9.1.7].

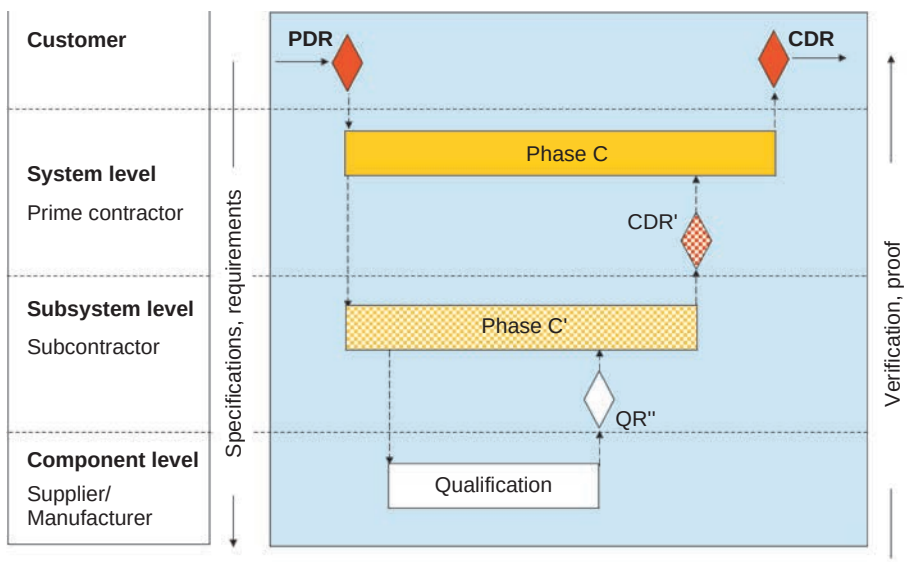


Figure 9.1.11: *Hierarchy of responsibility.*

9.1.2.3 Commercial Procurement Approach

Though development programs without strong economic pressure barely exist nowadays, the implementation of commercial programs differs from programs in which the development of methods, new techniques and scientific goals or political ambitions influence the risk and cost analyses.

Typical characteristics of commercial space programs are:

- High cost-effectiveness of the product (e.g., a satellite) through its use
- Low risks for development and operations.

They are the basis for deriving lower level system parameters and requirements:

- The system must be insurable against all negative influences in the satellite life cycle.
- Long operational lifetime.
- High reliability.
- Low cost for high utilization.

This implies for the manufacturer:

- Wide experience and high competence.
- Short development time and quick in-orbit delivery.
- All components and processes must be qualified and must have flight experience (heritage) to the maximum extent possible.
- Acceptance of high penalties in case of quality changes and delivery delays.

An additional characteristic of commercial programs is often the insufficient competence of customers regarding project specifics, particularly with respect to satellite technology. In order to maintain the needed balance between experienced contractors and their own management, customers often integrate independent experts (**consultants**) onto their team to **monitor** and **control** the program with them. Customers commonly understand their business as one of **service providers** offering satellite or system capabilities to their clienteles.

Therefore it is not surprising that in the customers' or clienteles' view the provision, operation and use of a complex satellite system is seen as a single procurement which is evaluated as an essential aspect of their envisaged business model.

The execution of a procurement program is conceptually different from the phased approach of ESA for development programs. It consists of five phases from the beginning (feasibility study) until system handover and use by the customer (see Figure 9.1.12).

Phase 1 Feasibility Study

- Investigation of existing technical know-how and market potential
- Development of a business model
- Development of technology solutions including boundary conditions
- Cost and economic concepts
- Evaluation of risk factors.

Phase 2 Request for Information Phase (RFI)

- Description of planned procurement by the customer
- Acknowledgment of interest in tendering by the bidder industry with
  - Preliminary cost and time estimates
  - Draft system concepts
  - Information about experience and heritage
- Development or validation of evaluation processes including software to be used during assessment and evaluations
- Evaluation of RFI from several bidders.

No.	Process title
1	Feasibility study
2	Request for information (RFP) phase
3	Request for proposal (RFP) phase
4	Production phase
5	Operations and maintenance
6	Disposal

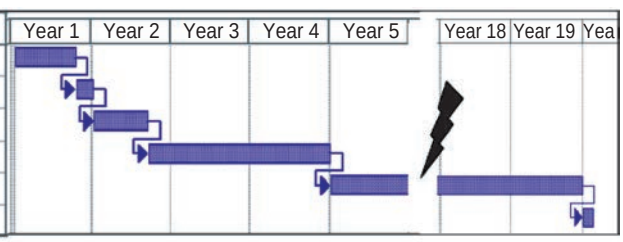


Figure 9.1.12: *Typical implementation flow for the procurement of a commercial satellite system.*

Phase 3 Request for Proposal (RFP) Process
See Section 9.1.6.1.

Phase 4 Production Process/Procurement Phase
- Manufacture of the contracted product by the main contractor
- Control and monitoring of the manufacturing processes by the customer to minimize project risks
- Acceptance and in-orbit delivery of the satellite
- Acceptance and commissioning of the system.

It should be noted that the manufacturing process is performed according to the contractor's internal procedures often based on ECSS standards.

Phase 5 Operations and Service Phase
- Transfer and operation of the complete system
- Operation and maintenance of the system to assure the required service for the customer and end users.

The customers of a procurement program acknowledge contract fulfillment only after successful in-orbit delivery, which can be either the satellite system which they themselves will operate, or the functional services of the system.

In the first case the results achieved (progress) by the contractor will be paid for by the customer during the implementation phase according to specified payment milestones. In the second case the contractor acts as a service provider and offers the satellite services after system delivery to the customer. The service is provided within the framework of a **service level agreement** (SLA) after commissioning of the system for a contractually agreed time.

The decision on which alternative is selected is driven mainly by economic considerations.

In this context the procurement process of the German Army's procurement offices for military goods and technologies, known as **customer product management** (CPM), should be mentioned. It is relevant because German military space programs follow this process, in particularly for:
- **SAR-Lupe**, a constellation of reconnaissance satellites
- **SATCOM Bw system**, consisting of two communications satellites.

The CPM approach (Figure 9.1.13) requires the manufacture of a prototype or demonstrator which permits successful verification of functional requirements.

9.1.3 Disciplines of Project Management

9.1.3.1 Planning

The focus of project planning is the development of a management plan or a project handbook (see Section 9.1.5.1 or 9.1.5.2). Based on the specific product tree, the established processes of the companies and the available resources and capabilities, the tasks, costs and schedules required to fulfill the mission goals must be planned. Project management must consider the main aspects described below.

Project Structure (Work Flow Plan)
The objective of a **work breakdown structure** (WBS) is to structure the project in reasonable and effective work packages to be distributed among the involved project partners, as well as to define the implementation phases or component groups. On the basis of the WBS, depending on the complexity of the program, concrete and reliable cost and schedule plans can be developed and integrated into the proposals.

In Figure 9.1.14 a typical WBS for a satellite program is shown. Each block represents a specific task (work package) to be performed by a responsible team member. Additional lower level structures may apply if clear and detailed interfaces to other tasks and team members are possible or necessary.

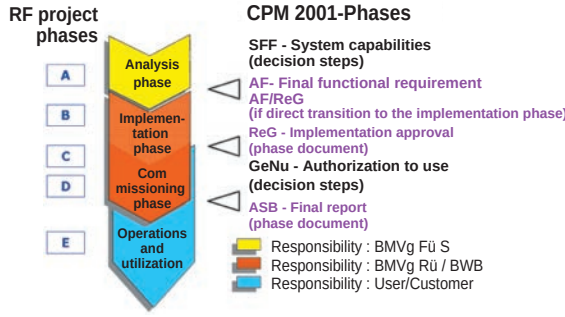


Figure 9.1.13: *Customer product management (CPM).*

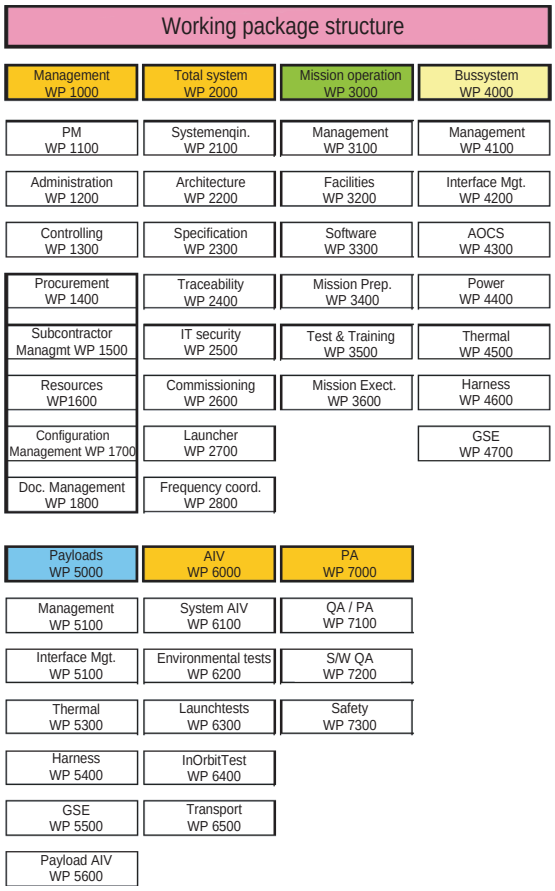


Figure 9.1.14: *Typical WBS structure for a satellite program.*

WP 1100		Issue Date : Sheet 1 of 1	
Project Title:	Title of Project		
WP Title:	Project Management		
Responsibility:	Project Manager		
Start Event:	Kick Off Meeting	**Planned Date:**	TBD
End Event:	End of Contract	**Planned Date:**	TBD

Main Objective:
- Assurance of excellent management of project tasks
- Coordination of work
- Key point to customer
- Assure customer satisfaction

Input Required:
- All project relevant data and information by customer
-

Tasks:
- Focal point to customer in all project relevant tasks
- Responsible for the project wrt technical, financial and programmatic aspects and tasks internal and against customer
- Change management
-

Output:
- Monthly reports
- Phase summary reports
- Minutes of Meeting protocol
- Contribution to deliverable documents

Figure 9.1.15: *Work package description.*

For a detailed definition and description of a task, a standardized WBS form should be used. Figure 9.1.15 shows the **work package description** (WPD) for a project manager as an example.

The quality of a WBS and a WPD is related to the level of detail and completeness as well as to the clarity of explanations provided, particularly for the following:

- The objectives of the work package.
- The needed preconditions or input data.
- The tasks to be performed.
- A detailed definition of the service or deliverable.

A bad and insufficient WPD may often result in disputes with customers or the project partners involved, since it normally has contractual implications.

Not least important, the quality of a WBS gives hints on the competence, experience and professionalism of the project management.

Time and Cost Planning

After the systematic structuring of all tasks needed to realize the product tree, the following steps are essential for completing the overall planning:

- Identification and collection of tasks (process steps) and their linkage according to the process logic or flow
- Definition of time for the implementation of each work item
- Identification of resources (responsibility) for each task
- Identification of cost elements per work item.

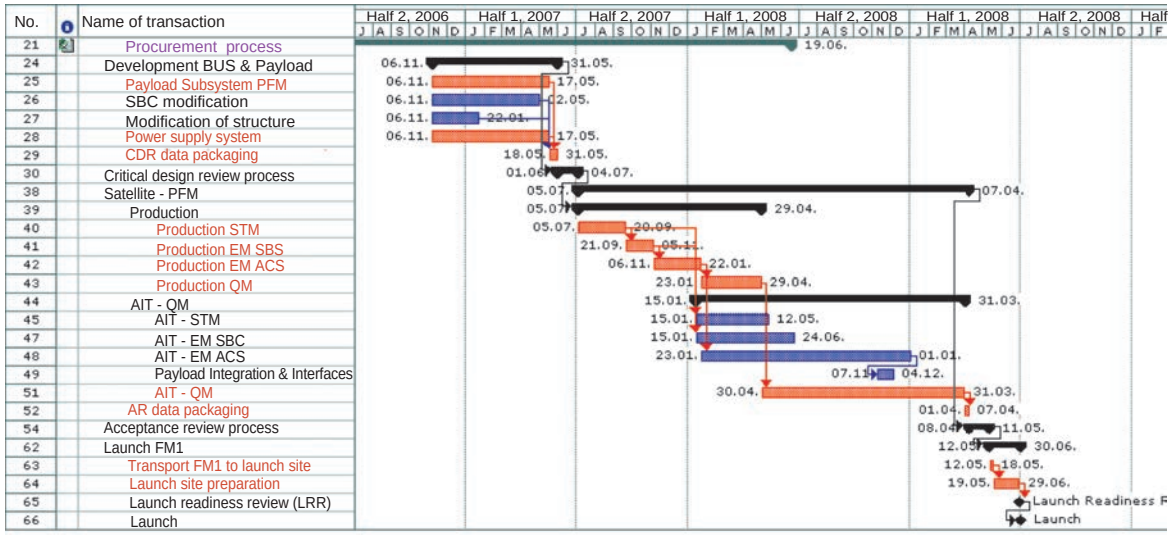


Figure 9.1.16: *Schedule.*

By using standard software tools, various graphics, diagrams and tables it is possible to visualize those critical and important project parameters for effective planning and control. Figure 9.1.16 shows a representative, partial, satellite program schedule. What is interesting here is the possibility to adjust single process steps to higher level process definitions and to identify the **critical path** of the program by specifying deadlines.

Figure 9.1.17 shows as an example the workload of a PM for the duration of a project. By consistently applying the plan and comparing it to the actual status, deviations and discrepancies can be identified early on and corrected.

The efficiency and value of the planning effort may be enhanced by consideration of further project requirements, such as:

- Definition of the results expected from the various reviews (see "Review Planning" below)
- Definition of milestones (see "Milestone Planning")
- Identification of documents
- Integration of lower level WBS (e.g., of subcontractors).

The so-called master plan represents the highest level of tasks and processes. It includes all delivery dates and serves as the basis for controlling the project. An important part of the master plan is the **assembly, integration and validation** (AIV) plan (see below).

Review Planning

Reviews allow coordination of the program by industry and also allow the customer to monitor progress of the project. Reviews are usually linked to the payment plan.

Contractors are obligated to invite customer representatives to all project-relevant reviews. Customers have the right to postpone reviews for justifiable reasons, for example the incomplete or delayed delivery of review documents (**review data package**).

Reviews complement the milestone plan and are used at all project levels and among all partners. Reviews may be seen as a **review process**, as indicated in Section 9.1.6.2.

Milestone Plan

Milestones are events within the AIV phase; they assure the quality of work (process) by comparison to appropriately predefined verification measures. A distinction is made between the following two points:

- **Key Inspection Point (KIP):** KIPs are for example defined for activities like checks and inspections performed by quality assurance representatives after completion of appropriate work or process steps.

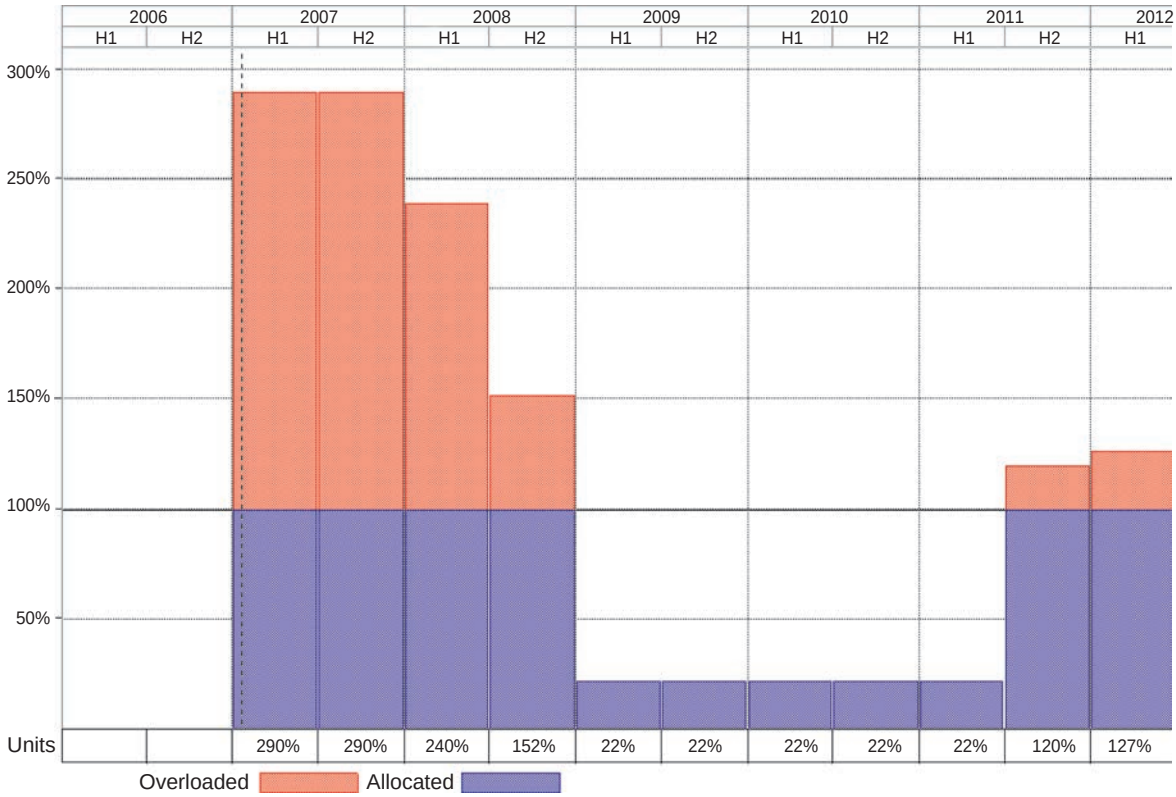


Figure 9.1.17: *Resource workload.*

- **Mandatory Inspection Point (MIP):** MIPs are for example defined for approval of critical activities with respect to the continuation of the AIV process. The customer's approval of the inspection results is mandatory as part of the obligation to provide support.

KIPs and MIPs are defined in the AIV plan together with the **product assurance** (PA) representative.

Acceptance Planning

Acceptance planning for a deliverable item or a service is based on contractual requirement documents or specifications and is supported by a **verification control document** (VCD) developed by the system engineer or project manager.

The VCD contains all technical specifications and definitions of how each requirement must be verified and traced. Typical verification methods are:

- Similarity
- Analysis
- Review of design
- Simulation
- Inspection
- Test.

The fulfillment of verification is documented and tracked by reference to the reports and results for each requirement or specification.

Additionally, contractors develop an **acceptance data package** (ADP) or **end item data package** (EIDP) for each item which will be delivered to the customer for assessment and evaluation. See also "Review Planning" above.

Table 9.1.3 shows a typical ADP for a space hardware product as part of a delivery from the space industry.

Table 9.1.3: *Acceptance data package (ADP).*

COVER SHEET AND TABLE OF CONTENTS
CHANGE RECORD AND DISTRIBUTION LIST
STATEMENT OF COMPLIANCE
DELIVERED ITEMS DATA:
• LIST OF DELIVERED ITEMS
• PACKAGING/HANDLING TRANSPORT AND STORAGE DATA SHEET
• SHIPPING DOCUMENTS
CERTIFICATES
C.I.D.L./AS BUILT CONFIGURATION LIST
NOTES AND COMMENTS
HISTORICAL RECORD
OPERATING TIME/CYCLES RECORD
CONNECTOR MATE-DEMATE RECORD
AGE/LIMITED LIFE ITEMS:
• LIMITED LIFE ITEMS LIST
• LIMITED LIFE RECORD
• AGE SENSITIVE ITEMS RECORD
MISSING ITEMS LIST
TEMPORARILY INSTALLED ITEMS RECORD
LOOSE DELIVERED ITEMS LIST
NON CONFORMANCE LIST AND COPIES OF ALL MAJOR NCRs
RFW LIST AND COPIES OF ALL RFWs
RFD LIST AND COPIES OF ALL RFDs
ECP/CCN LIST AND COPIES OF ALL ECPs/CCNs
NON STANDARD CALIBRATION ITEMS LIST
SAFETY DATA PACKAGE
TEST PLAN
PROCEDURES
TEST REPORTS
INSPECTION REPORTS
OPEN WORK/OPEN TEST/DEFERRED WORK
MASS AND POWER BUDGET RECORDS:
MASS BUDGET RECORD SHEET
ELECTRICAL CONSUMPTION BUDGET RECORD SHEET
OPERATION AND MAINTENANCE MANUAL
TOP ASSEMBLY DRAWINGS, INTERFACE DRAWINGS
PHOTOS
PACKING, UNPACKING, HANDLING, STORAGE AND TRANSPORTATION PROCEDURES
PRESSURE VESSEL DATA:
• LIMITED LIFE PRESSURE VESSEL ITEMS LIST
• PRESSURE OPERATING TIME/CYCLES RECORD
PYROTECHNIC, RADIOACTIVE AND LASER ITEMS DATA LIST
TRR/QR/AR/DR MINUTES
INTERFACE DESCRIPTION
DESIGN DESCRIPTION

9.1.3.2 Project Leadership

Though not the focus of this chapter, leadership is essential and often the cause of problems and failures if not executed correctly.

In particular, large projects represent small, autonomously organized units whose team members have a clearly defined goal and tasks and a recognized identity distinguishing them from "the rest" for a period of time.

A project manager's claim to leadership for executing a project under his or her sole responsibility is legitimized by an announcement and assignment of obligations and authority by the management of the organization and is based on the mission and project goals.

Clearly, the following personal characteristics and behavior of a project manager contribute to a project's success:

- Identification with the mission objectives and project goals.
- Authority
- Resoluteness
- Capability of teamwork
- Capability of motivating
- Capability of moderating
- Loyalty
- Fairness and transparency
- Communication behavior
- Abstraction and analytical capabilities
- Verbal and rhetorical skills
- Capability of self-criticism and learning
- Resilience.

Which of these capabilities will positively affect the project depend on the kind of project, its criticality, size and complexity, as well as on the team members and institutions involved.

Loss of trust by the project partners caused by insufficient capabilities is often attributed to "bad management" and is difficult to regain.

9.1.3.3 Project Control

An essential task of project management is monitoring project performance.

The objective is to determine the actual project parameters and factors in correlation with planned values, as well as to analyze deviations and consequences by defining preventive and corrective measures.

The most important elements of this assessment are:

- Cost elements
- Technical approach
- Requested quality
- Time and implementation schedule.

The results of assessments are summarized in an internal status report delivered to the internal management or, if requested and contractually agreed, to the customer.

9.1.3.4 Communication and Reporting

Communication between project team members is an essential instrument of effective and successful project management (see Figure 9.1.1). The content and amount of the needed communication effort is related to the tasks and interests of the project partners involved (see Table 9.1.1).

By no means should this management task be underestimated, because bad communication generates mistrust and dissatisfaction and always results in bad project relationships.

The portfolio of successful project management includes the following tasks.

Interfaces to the Project Team:

- Regular team meetings
- Up-to-date and complete exchange of information
- Positive and negative feedback
- Formulation of problems
- Clear task requirements
- Clear and unambiguous work package descriptions
- Efficient workloads which avoid continuous overburdening.

Interfaces to Internal Management: Regular and formalized project reviews are recommended, informing about:

- Resource loads
- Capacity
- Risks
- Problems
- Quality
- Cost
- Schedule.

Status reports supplied on a weekly or monthly basis depending on the size of the project or the risks involved are common.

Interface to the Customer: If not already contractually agreed, the contractor's project management should report on a regular basis to the customer's representatives. The following information should be provided as a minimum:

- Project progress, including the most important events of the reporting period
- Planned activities for the next reporting period
- Agreed deviations or discrepancies from functional requirements and specifications
- Identified problem areas and risks
- Status of quality
- Actual schedule, including work accomplished to date
- Status of cash flow and planning
- Technical budgets, control and schedules.

Furthermore, verbal communications with the customer's representatives on a regular basis are recommended, as is the provision of ad-hoc information in case of critical events.

Additionally, it is recommended that all data and information is stored on a jointly accessible project server to allow the customer to carry out ad-hoc reviews and evaluation of documentation. As an example, consider the master schedule. This schedule is discussed only at the highest level due to its extraordinary size in the case of large programs. However, a detailed and complete schedule covering all working levels must be updated at appropriate intervals and always available to facilitate transparent communication.

9.1.4 Tools of Project Management

9.1.4.1 Organizational Structures

Companies and organizations with an increasing number of employees may be structured according to their specific areas of competence and functions. Figure 9.1.18 shows a typical example of a **line organization**.

Due to the interdisciplinary character of projects it may be difficult to establish and guide project teams. Cooperation and control may be difficult, particularly in cases of conflict since the decision path and process are complex and time consuming.

Such structures may lead to inefficient bureaucracy and are not useful for large-scale projects. For programs requiring flexible and systematic structures, a **matrix organization** is recommended (Figure 9.1.19).

In this organizational form the management establishes a project team under the control of the leading project manager. The delegation of overall responsibility with a direct reporting path to the management is essential.

With this delegated responsibility the project manager can establish and control a specific project's organization adapted to the program goals. An example of this for a satellite project is shown in Figure 9.1.20.

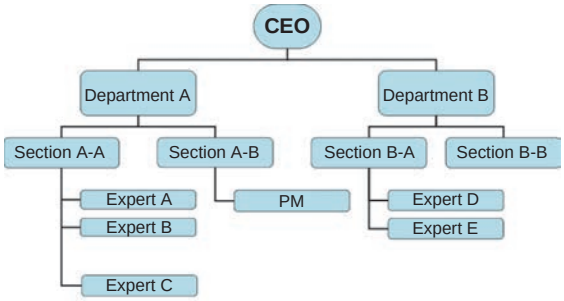


Figure 9.1.18: ***Line organization.***

9.1.4.2 Model Philosophy

The model philosophy is developed from the mission and program requirements. For programs following the ESA approach the model philosophy is developed in Phase A and finalized in Phase B.

The variances are based on value for money, quality and risk considerations, and define the kind of models to be developed, including the depth of verification and test spectrum. See also Figure 9.1.16 and Section 8.3.2.

Although the model philosophy is developed by technical engineers, the project manager supports the decision process for selecting an optimized solution. Only when the types and number of models and processes are agreed can they be considered in the implementation planning.

Procurement programs with minimum development risks produce only **flight models** for cost reasons and because they can rely on the long experience of manufacturers. Qualification models or prototypes are usually not considered. However, it is not unusual that even in flight models components or items are used which are not available as commercial off-the-shelf(COTS) products and which have to be specially developed for the program.

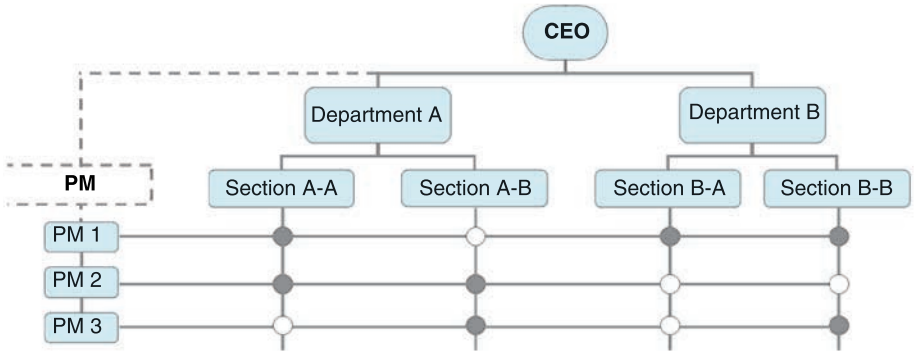


Figure 9.1.19: ***Matrix organization.***

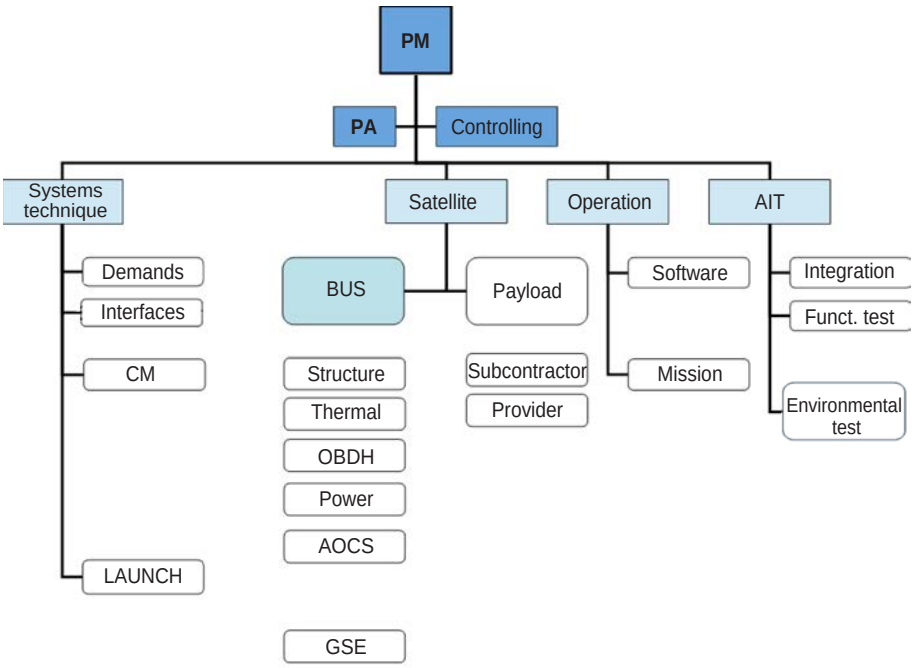


Figure 9.1.20: *Satellite project organization (example).*

9.1.4.3 Risk Management

Risk management must identify and eliminate risks as early as possible with appropriate preventive measures to avoid or limit possible damage which may materialize during the program. Thus the focus must be on risks which may affect the success of the mission or project. Additionally, risk management supports the achievement of realistic schedule, cost, and resource planning for the program.

Risk must be identified and included in a **risk matrix** with keyword information on the following items:

- **Risk Item Description:** Description of the risk with respect to cost, time and quality requirements.
- **Cause of Risk:** Summary of all possible causes for the risk.
- **Description of Consequence:** Summary of consequences and impact on project success and costs.
- **Probability of Risk (1–100%):** Estimation of probability of risk occurrence.
- **Effect of Risk (1–10):** Evaluation of the risk effect based on predefined criteria.
- **Criticality of Risk:** Calculated by multiplying the risk probability by the effect of the risk.
- **Recovery Indicator (time in months):** Estimate of time within which the risk effect can be corrected, for calculating repair costs.
- **Risk Owner:** Designation by the PM of the person responsible for managing the risk.
- **Technical Mitigation:** Description of preventive technical measures for avoiding risk.
- **Contractual Mitigation:** Description of contractual measures to control risks.
- **Contingency:** Description of measures which have to be initiated if the risk occurs.

All the defined resulting measures must be monitored on a regular basis by project management and their status reported to and discussed with the internal management.

9.1.4.4 Cost and Schedule Planning

The main contractor must work out an **overall schedule** for program monitoring and control in which all relevant measures, work and process steps and milestones are included. In this sense the schedule may be seen as the **transfer of work package descriptions into processes**; these also include all the tasks assigned to subcontractors.

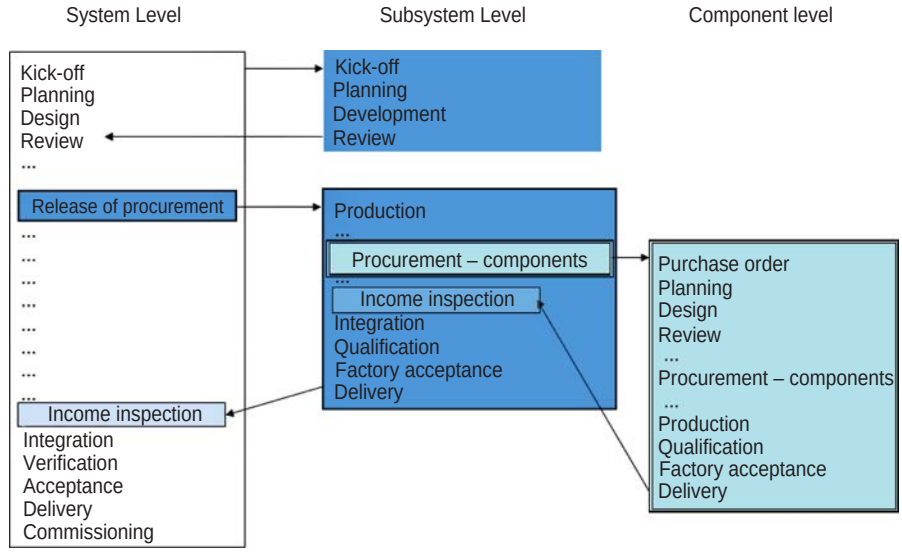


Figure 9.1.21: *Systematics of the schedule.*

For effective project control each process must provide the following information:

- Task description of the individual process steps
- Work package number
- Responsible partner (name and organization)
- Identification of milestones (review, MIP, KIP)
- Cost elements (fixed costs, variable costs, travel costs)
- Time duration with date of project start and fixed limitations
- Work effort
- Resource planning
- Documents to be provided on completion of the work package (if applicable).

The following aspects are therefore essential:

- All subsequent **process steps** must follow a closed loop concept and a standard structure with the following contents:
  - Receiving tasks (e.g., KIP, inspection)
  - Performing tasks
  - Approval, checking task results
  - Delivery/handover to the next subprocess.
- **Linking** of all relevant processes according to the workflow.

Only if the above requirements are fulfilled can the **planning tasks** be efficiently performed. These are:

- Establishment and maintenance of a master plan
- Analysis of the critical path
- Presentation of project reports (cost summary, resource capacity and consumption, status)
- Presentation of Gantt diagrams
- Presentation of Pert diagrams
- Change management.

The use of commonly available **software tools** for project planning and control is recommended.

Because of the importance of project planning, the schedule for the project should be available to all project partners on a common server.

Due to the increasing **complexity** of large-scale space programs, thousands of individual process steps may be necessary to achieve the mission goal. In this case it is essential to subdivide the overall project into smaller projects and to link them accordingly (see Figure 9.1.21).

To better manage the schedule and assure its functionality, the following aspects should be the basis for its implementation:

- Clarity of single (sub)processes
- Similar level of detail for each process
- Product/process-oriented structure of the schedule
- Integration of internal milestones as KIP
- Integration of subcontractor and supplier milestones
- Color coding of essential process steps

- Linking of individual work steps within larger blocks; minimization of links across blocks.

Maintenance of the schedule should be carried out and coordinated by project management to assure in-time control and monitoring of the progress of work.

The **synchronization and updating intervals** should be shortened as the project progresses, from two months at the beginning of the project down to a weekly update sequence for critical activities. The level of detail should increase accordingly.

9.1.4.5 Quality Management Aspects

Quality management plays both a preventive and a corrective role in achieving high project and product quality; the goal is reliable and efficient project performance to reach the mission aims.

International standards for space applications (**ECSS**), industrial applications (**ISO**) and **in-house rules** are applicable. The application of in-house standards is meant to ensure that the optimized, qualified and certified processes of the companies involved in the project are used to achieve the program goals.

Quality assurance standards may be tailored to technical and programmatic requirements, cost and time demands and should take the experience of former projects (heritage) into account.

It is the task of the management to establish an efficient **quality assurance** or **product assurance** process which is independent of the engineering process and reports directly to the project management.

9.1.4.6 Configuration Management

The goal of configuration management is to document the status of delivered services, functions and physical requirements, providing full transparency. Software, hardware, drawings and documents are to be documented and archived, in each case with the revision status, so that an overview of the "as-built" status is available at any time.

Configuration management documents the product tree and identifies all hardware and software in the **configuration item data list** (CIDL).

Configuration management may be seen as a supporting process, controlled by the quality management systems of the organization (e.g., according to ISO 9001 and ISO EN 9100).

9.1.4.7 Logistic Aspects

In order to support economical services or operations with minimum life cycle costs, the basics of **integrated logistic support** (ILS) and **logistic support analysis** (LSA) should be applied early in the operational phases, in particular to support efficient **interoperability** with other missions (multimission compatibility).

The number of items to be controlled is based on the product tree and must be included in the cost plan. The most important cost items include:

- Space segment (development and procurement):
  - Satellite platform
  - Payload compound
  - Test and analyses
  - Orbit transfer
- Ground segment:
  - Infrastructure (leasing rates)
  - Technical equipment (leasing rates and purchasing)
  - Operation (including leasing rates)
  - Ground support equipment
- Program management, for example:
  - Project and quality management
  - Procurement and purchasing
- Disposal of all products
- Personnel cost and training/qualification
- Maintenance and spare parts
- Changes and modifications to hardware and software.

In particular, requirements concerning **mean time for repair** (MTFR) must be analyzed and evaluated appropriately.

Business models for selected components and services of the space segment as well as the ground segment (including personnel, training and qualification) require a standardized **ILS concept** for mission operation. This should provide information on:

- Logistic requirements for construction, reliability, material stability

- Maintenance and management of materials and tools (overall concept, stages as required, equipment pools)
- Software maintenance and change
- Storage, conservation, packing and labeling, in particularly for long lead items
- Cataloging and indexing
- Training and training tools.

9.1.4.8 IT Security Aspects

The function of IT security is to handle, store and communicate information in such a way that its security (confidentiality), integrity, availability and authenticity are assured.

Security of data means that information about and data resulting from operations and services are only provided to authorized persons or organizations.

Integrity means the protection of data against manipulation and disturbance.

To assure data security, various measures must be taken which can be categorized into four groups and are described in the following.

Technical Measures:
- Safe locking of the satellite transport container for transport to the launch pad
- Encryption of data transfer and transmission
- Safe storage on secured data storage media
- Installation and operation of firewalls for protection against hacking, data manipulation and deletion
- Use of crypto units
- Authentication by passwords.

Organizational Measures:
- Definition of responsibilities and authorization of ground segment operations personnel
- Complete documentation control of corrective, installation and maintenance work on hardware and software.

Infrastructural Measures:
- Establishment of secure areas including access control adapted to specific operational phases.

Personnel Measures:
- Security checks of personnel
- Agreements with the personnel involved to ensure that data classified as *confidential* is not transferred to unauthorized persons.

The effort in carrying out the measures briefly described above depends on the requirements and size of the project. On request they may be analyzed and defined in a security concept which is integrated into the program by project management.

9.1.4.9 Safety Aspects

National and international laws and regulations require specific consideration of safety-relevant aspects during realization of a project.

Safety is primarily related to **human safety**, that is to prevent injury or death, but it is also related to **damage to satellites** and lower level subsystems and components.

Safety aspects must be analyzed systematically and considered throughout the entire AIT process as well as during the life cycle of a satellite.

The mission safety program consists of the following activities as a minimum:
- Definition of safety-relevant requirements in the context of a **product assurance and safety** (PA&S) plan
- Performance of a tailored **safety analysis**.

These measures take into account the complete product tree, in particular work on the spacecraft and use of **ground support equipment** (GSE) as well as transport and lifting tools.

Work at the launch pad is controlled by the relevant site safety procedures.

The safety analysis is performed at the beginning of the program with the aim of identifying safety-critical items and considering appropriate measures early in the design so that the relevant processes are covered.

Typical safety-relevant aspects are:
- Collision
- Contamination
- Electric shock
- Explosion
- Fire
- Injury and illness
- Temperature and pressure extremes

- Falls
- Lacerations.

On the basis of the safety analysis and the identified critical areas, specific safety procedures are developed and monitored by the safety engineer.

9.1.5 Project Management Documentation

9.1.5.1 Management Plan

The management plan is an element of the design and planning work (see Figure 9.1.22), and has to be established in Phase A of a space program according to ESA requirements.

The phase of determining concepts and specifications leads to the establishment of a management plan which defines and describes the structural and methodical approaches and the resulting processes required to successfully meet the mission goals with respect to technical realization, time plan and cost framework.

The management plan is structured in such a way that, as the program proceeds, modifications can be easily made or incorporated into the project handbook (see Section 9.1.5.2), up until kick-off of the implementation phase (Phases C and D). At this milestone the management plan must be finalized and subject to document control.

Table 9.1.4 shows the structure of a management plan.

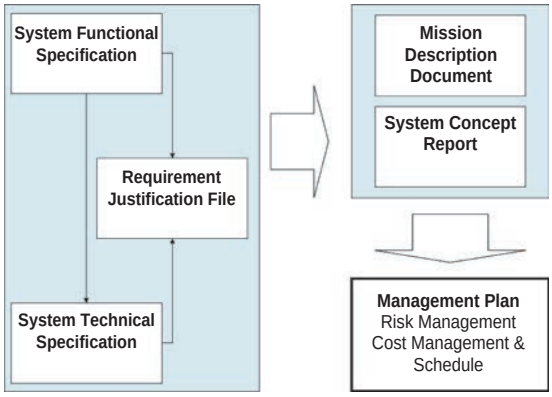


Figure 9.1.22: *Phase A planning documentation.*

Table 9.1.4: *Structure of a management plan.*

1	**Introduction**
1.1	General
1.2	Objectives
1.3	Applicability
1.4	Scope
1.5	Compliance
2	**Documents**
2.1	Applicable documents
2.2	Reference documents
2.3	Normative documents
3	**Definitions**
4	**Mission description**
4.1	General
4.2	Mission goals
4.3	Overview
4.4	Technical system description
4.4.1	Satellite system
4.4.2	Platform
4.4.3	Payload
4.4.5	Functional tree
4.4.6	Product tree
5	**Risk management**
5.1	Introduction
5.2	Identified risks
6	**Project organization and management**
6.1	Project organization and tasks
6.2	Project management
6.3	Work breakdown planning
6.4	MIP/KIP planning
6.5	Project planning and reviews
6.6	System acceptance
6.6.1	General
6.6.2	Acceptance criteria
6.6.3	Non-compliance of requirements and goals
6.6.4	Final acceptance
6.7	Project management tools
7	**Quality management**
7.1	Configuration management
7.2	Documentation management

Table 9.1.4: Continued

7.3	Reporting
7.3.1	Status report
7.3.2	Final report
8	**Schedule planning**
8.1	General
8.2	Planning approach
8.3	Structure of master schedule
9	**Cost planning**
9.1	General
9.2	Definition and acronyms
9.3	Calculation and cost planning
9.4	Cost control and reporting

An essential part of the planning activities is the characterization and summary of the deliverable documentation. Table 9.1.5 shows as an example an (incomplete) list of deliverable documents with typical delivery milestones. The final and binding list must be defined within the framework of management planning at the end of Phase B (approved at the PRR) according to the standard ESA project management approach.

9.1.5.2 Project Handbook

A project handbook should be drawn up and issued by the program's contractor, particularly if detailed management planning as described in Section 9.1.5.1 is not available or if standardized procurement programs are not realized according to ESA guidelines.

The aims and objectives of a project handbook are to provide a complete description of the responsibilities, roles and tasks of all the partners involved as well as a definition of the communication, control and acceptance processes. Table 9.1.6 shows a typical structure for a project handbook.

It should be mentioned that elaboration of a management plan or a project handbook is onerous because it is a time-consuming task; its value, however, becomes evident with time. A well-structured and thoroughly defined planning process supports the overall understanding of all the project partners

Table 9.1.5: *List of deliverable documents according to ESA.*

Deliverable documentation	MDR	PRR	SRR	PDR	CDR
System functional specification	D	F			
System technical specification	D	F			
Justification file	D	F			
System concept report	F				
Mission description document	D	F			
Management plan	**D**	**F**	**U**	**U**	
Technical specification		D	F		
PA plan			D	F	U
Design definition file		D	F		
Design justification file		D	F		
Interface requirements document			D	F	
Configuration management plan			D	F	
Document management procedure			D	F	
Analyses				D	F
Specifications					F
AIT plan				D	F

Documentation status: D = Draft, F = Final, U = Update.

Table 9.1.6: *Structure of a project handbook.*

1	**General**
1.1	Introduction
1.2	Objectives
1.3	Scope
1.4	Classification
2	**Documents**
2.1	Applicable documents
2.2	Reference documents
2.3	Normative documents
3	**Project management**
3.1	Responsibilities
3.1.1	Functions
3.1.2	Work packages
3.2	Communication
3.3	Management meetings
3.4	Interfaces
3.5	Project schedule
3.6	Risk management
3.6.1	Technical risks
3.6.2	Programmatic risks
4	**Quality management**
4.1	Documentation management
4.1.1	Type of documents
4.1.2	Document approval and release
4.1.3	Archiving of documents
4.2	Controlling
4.2.1	Controlling methods
4.2.2	Controlling tool
4.3	Function of external consultancy
4.4	Reporting
4.4.1	Status report
4.4.2	Final report
4.5	Non-conformance management
4.5.1	Criteria of non-conformances
4.5.2	Escalation principles
5	**Acceptance**
5.1	Break milestones
5.2	System acceptance

and essentially supports the establishment of the necessary trust needed for the successful performance of the project.

9.1.6 Customer–Contractor Relations

A successful project is characterized by positive relations between the customer and contractor. This essentially important partnership is based on trust as well as a common understanding of the requirements and core processes.

9.1.6.1 Procurement Process

The procurement process (Phase 3, Section 9.1.2.3) is initiated after the evaluation of the RFI phase and consists of the following steps:

1. Generation of procurement documents (RFP, see Figure 9.1.23) by the customer, assisted by independent experts (**consultants**).
2. Submission of proposals by industry based on the RFP documentation.
3. Evaluation of the proposals by the customer with support from the consultant(s).
4. Selection of the best offer following a predefined evaluation and selection process.
5. Contract negotiations with the selected contractor.

Contracts between the customer and contractor are established on the basis of the RFP documentation and most likely specify a **firm fixed price**. In the case of

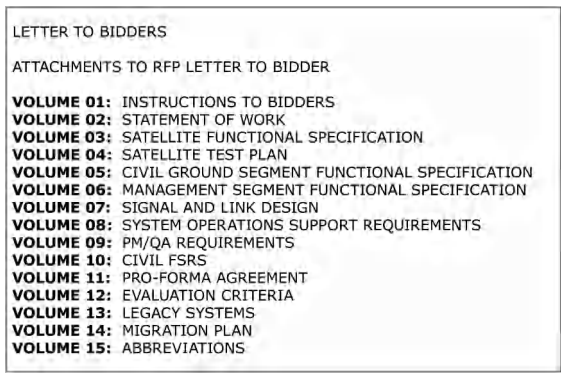


Figure 9.1.23: *RFP for a procurement process.*

service contracts, so-called "service level agreements" (SLAs) may be contracted with specific penalties if the contracted services are not, or are only incompletely, provided. Operational risks are covered by adequate insurance.

9.1.6.2 Review Process

Reviews may be seen as essential management tasks for controlling the services delivered by the contractor and for checking deliveries in the context of a review board chaired by the customer or the customer's representative. The purpose of a review board is to approve or accept these services or products (see Section 9.1.6.3) in relation to the requirements on the basis of the delivered documentation.

In the case of positive confirmation, the past phase is considered to have been successfully completed and authorization to proceed is granted (see Figure 9.1.24). Identified discrepancies or recommendations from the review board are documented in **review identified discrepancy** (RID) reports, which are submitted to the contractor for further consideration.

Projects controlled by ESA regulations are distinguished by a systematic and extensive review process consisting of the following process steps:

1. Delivery of the **review data package** to the customer by the contractor.
2. Comprehensive review of delivered documentation (2–3 weeks) by the customer.
3. Presentation of review data package at the customer's premises by the contractor's personnel.
4. Identification of RIDs by the customer's review team (7–10 days).
5. Response to the RIDs by the contractor's team (5–10 days).
6. Presentation of RIDs at the customer's premises (2 days) by the customer.
7. Summary assessment and evaluation of the review data package by the customer.

Consistently following this process requires about 4–5 weeks to formal confirmation, when authorization to proceed to the next phase is provided by the customer.

For space programs with a more economical structure the review process is reduced accordingly:

1. Delivery of the review data package to the customer by the contractor.
2. Comprehensive review of delivered documentation (2 weeks) including issue of RIDs by the customer.
3. Presentation of the review data package and the project status at the customer's premises, including the generation of RIDs.
4. Improvement of the review data package and resolution of the RIDs by the contractor (5–10 days).

Continuation of the project is decided at the review under the condition that all RIDs are closed out within the identified time frame.

Review identified discrepancy (RID)

- Identification of review item
- Reference to relevant documentation
- Description of discrepancy
- Description of effect
- Recommendation
- Description of measures (actions for improvements) by the contractor
- Evaluation of defined activities
- Final determination that all defined measures have been taken.

All review dates are contractually defined and to be confirmed by the contract partners at least four weeks prior to a review. The agenda and the list of participants must be distributed to the review participants at least two weeks in advance so that any changes requested by the partners can be considered.

Reviews must be documented "online" by an appropriate meeting protocol, signed by the representatives of the parties involved. Meeting protocols may be used as contractual documents in cases of conflict.

9.1.6.3 Acceptance

As a baseline, acceptance must take place in reviews. For smaller projects the partners sometimes omit the review for time reasons, but this is often

accompanied by unsystematic and insufficient execution of tasks.

Acceptance in the context of a review process can only be successful and traceable if the aims and objectives of acceptance including the acceptance criteria are unambiguously defined. If not already contractually defined prior to the review, the success criteria must be defined by the customer. See also Figure 9.1.24.

If the review is successful, the past project phase is concluded and the respective payment for the accepted service or delivery is initiated.

By acceptance the customer confirms responsibility that the subject of acceptance complies with the requirements. The contractor in turn confirms that the delivered service or item complies with the contractual requirements.

Acceptance Criteria

Acceptance vis-à-vis the technical specification is documented in a **verification control document** (VCD) which is maintained and kept up to date throughout the implementation phase.

The VCD lists all the specifications and requirements. It also describes how and at which milestone the specified parameters must be verified.

Other specific criteria such as delivery time, status of **nonconformance reports** (NCRs), and waivers are contractually regulated and must be respected. Deviations from the acceptance criteria must be documented in a **contract change note** (CCN) and submitted to the customer for approval.

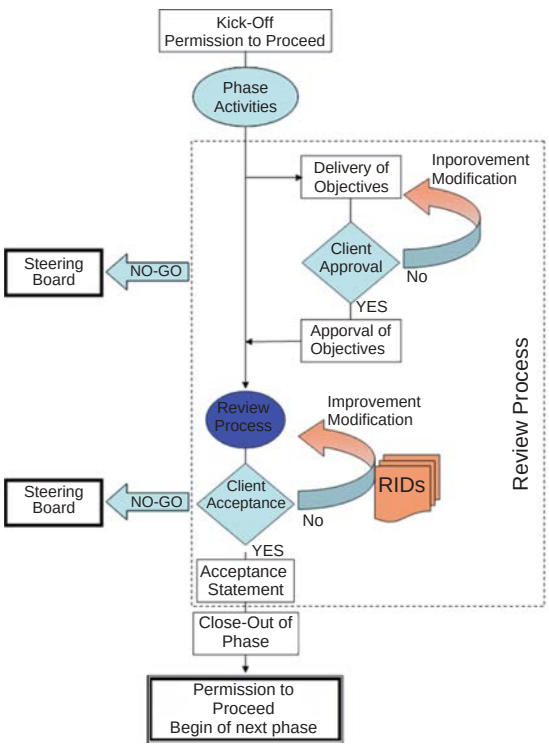


Figure 9.1.24: *Acceptance strategy.*

Nonconformance

Termination of a process, activity, test or operation may be initiated with the approval of those with the responsibility, or of instances if an identified risk might have a significant impact or even cause damage with respect to the following:

- Mission objective
- Project goals
- Cost and schedule planning
- Contract
- Environment
- Health and human life.

For each such event a nonconformance report must be issued. Any relevant contractual consequences must be documented.

9.1.6.4 Final Acceptance

Depending on the status of the project and ownership relations, final acceptance takes place either after successful commissioning or with the end of operations, if operations were part of the delivery.

Projects may be considered as being accepted if:

- Contractual work has been completely carried out and in the correct manner.
- Specifications and mission requirements have been fulfilled.
- All contractually defined deliverables (e.g., documentation, test results, drawings) have been delivered. Noncompliance with contractual requirements is regulated by the contract itself (e.g., in contract change notes).
- The necessary reviews and inspections have been executed without any discrepancies.

9.1.7 Conclusions

A critical analysis of projects, in particular of completed projects, is often presented in the form of a **lessons learned** report. This report can reveal whether:

- The project goals and objectives were partially missed.
- The cost and time budgets were exceeded.
- The quality was insufficient.
- The "team players" were completely overworked.
- The customer was dissatisfied.

These shortcomings can be avoided by systematic and goal-oriented work performance and by consistent and customer-oriented management. Project management is successful if the goals and objectives with respect to technical specifications, as well as to costs and time schedules, are met and recognized by the customer as being successful. This requires the establishment of management processes using suitable methods of planning, organization, control, results evaluation and analyses, and the immediate initiation of corrective measures or changes to achieve the overall goals of the project.

Thus, technical and methodological skills, but in particular also the social competence of the project manager, have an essential role:

- Project management is more often apt to fail because of erroneous decisions than because of methodological shortcomings.
- Unfortunately, failures are persistently repeated, followed by a firm resolution to avoid them in the next project.

While technical competence often establishes the preconditions for assigning project management responsibilities to candidates, less attention is paid to their methodological and social competences. The various interdependencies and consequences of successful project management have been outlined and described in this section.

Bibliography

[9.1.1] Wikipedia, freie Enzyklopädie. Schlagwort-Suchbegriff, December 2006.

[9.1.2] Zollondoz, H.D. Lexikon Qualitätsmanagement. *Handbuch des Modernen Managements auf der Basis des Qualitätsmanagements.* Munich: R. Oldenburg Verlag, 2001.

[9.1.3] Saynisch, M. *Konfigurationsmanagement.* Düsseldorf: Verlag TÜV Rheinland, 1984.

[9.1.4] Madauss, B. *Handbuch Projektmanagement*, 4. Auflage. Stuttgart: Verlag C. E. Poeschel, 1991.

[9.1.5] ECSS-E-10-02A. *Space Engineering, Verification.* November 1998.

[9.1.6] ECSS-M-30A. *Space Project Management, Project phasing and planning.* April 19, 1996.

[9.1.7] ECSS-M-00-02A. *Space Project Management, Tailoring of space standards.* April 25, 2000.

[9.1.8] Customer Product Management (CPM). Verfahrensbestimmungen für die Bedarfsermittlung und Bedarfsdeckung in der Bundeswehr, Erlass, May 24, 2004.

9.2 Quality Management

Jürgen Mathes

In this chapter the focus is not on general quality management according to the rules and standards of EN ISO 9001 or EN ISO 9100, but rather the subsequent processes are described for an adequate **quality management for space products** in praxis.

The prerequisite for the design, development, manufacture, integration and testing of space products is a **certified and implemented** generic quality management system within the relevant organization. Quality management looks after the management and main processes within the organization and is responsible for the implementation and verification of customer requirements and the quality policy for all products and projects. It must have sufficient independence to perform this control function both neutrally and objectively. This can be achieved by organizational implementation as a cross-functional responsibility within the organizational structure and by having quality management reports sent directly to the head of the organization.

Quality management must have the goals of **preventing** any product failures, optimizing the organization's processes, and providing the long-term product quality which assures customer satisfaction.

Quality cannot be implemented by decree, but is the result of its implementation by all employees involved in product generation. A prerequisite is a high sense of responsibility on the part of all employees in carrying out and evaluating their work and in assuring the continuous implementation of corrective measures.

The **effectiveness** of a quality management system requires an analysis of how processes are organized. When it is established it is essential that information flow, staff involvement, interrelationships and working conditions be clearly identified. Afterward, the organizational structure, functional and disciplinary responsibilities, and the responsibilities of individual employees, can be defined.

The terms, tasks and responsibilities referred to in this section are defined and discussed below.

9.2.1 Terms

Quality management, quality assurance and product assurance are terms which are often used in parallel, but they describe different tasks.

Quality Management

Quality management (QM) implements the generic and obligatory ISO quality standards in the **organization's processes**, and maintains its quality policy, manuals, procedures, facilities and resources.

Quality management is responsible for the following:

- **Translation of the organization's quality policy** into projects and products
- **Checking its processes**, manuals and procedures
- **Certification and qualification** of its facilities and personnel
- Definition of **parameters** and evaluation of **customer satisfaction**
- **Auditing of suppliers**, supplier ratings and quality acknowledgment
- **Internal auditing** and maintenance of technical **process qualifications**
- **Maintaining the quality** of equipment and facilities
- Supply of norms and standards
- Establishment of **quality control loops**
- Implementation of **risk management** related to quality.

Quality Assurance

Quality assurance (QA) is the operational and controlling function of quality management. It converts the rules of quality management into the **product-specific processes needed for product generation**. Quality assurance monitors all hardware and software life cycles during product generation. This includes assuring the traceability of design development, production and qualification of a product, including the associated product documentation, up until formal acceptance and delivery to the customer.

Other tasks concern the selection and approval of manufacturers and suppliers, the handling of deviations and defects, as well as failure and change management.

Quality assurance is responsible for the following:

- Planning **quality assurance measures** for the product generation process
- Definition of and adherence to product **quality standards**
- Transfer of quality requirements into **specifications and work statements**
- Preparation of product-associated **quality plans and work instructions**
- **Availability** of product-associated **facilities and equipment**
- **Verification** of product-associated customer requirements
- Implementation of **quality controls**, acceptances and certification
- Preparation and maintenance of the product-associated **quality documentation**.

Quality assurance follows a generic task sequence, as below.

Planning of Quality

Quality planning includes the definition of quality targets for a given product and the measures for their permanent monitoring (analytical QA). This planning is supplemented by preventive measures which assure the maintenance of quality (constructive QA). This planning requires the following:

- Depiction and consequences of product-specific processes
- Specification of sequence and interaction of processes
- Monitoring, measuring and analysis of processes
- Measures for permanent improvement of processes
- Assuring resources and information.

Assurance of Quality
Product quality is measured using defined test results or indicators. Analysis of these measurements leads to improvements in the constructive QA.

Guidance of Quality
The guidance of quality regulates the provisioning of resources (employees and material) on the basis of planning and quality measurements.

Control of Quality
Quality is controlled using uniform evaluation criteria based on the evaluation of indicators (see Section 9.2.2.3). A prerequisite is the formalization of process procedures, documents and document templates. This is not to be understood as bureaucratization, but as a measure which guarantees the objectivity of the quality assurance processes.

Product Assurance
Quality assurance for space products is referred to as product assurance (PA).

Product assurance translates generic quality assurance into **space-specific processes needed for product generation.** These processes are oriented to the rules and regulations of customer organizations such as ESA, NASA, DLR, as well as those of launcher authorities and user organizations such as Arianespace, COSMOS, METEOSAT, SES and ASTRA. Product assurance supports project management in carrying out the product life cycle and supervising product generation according to the customer's programmatic and technical requirements, and based on experience and lessons learned.

The tasks of space-specific product assurance are beyond the generic quality assurance tasks and therefore require a dedicated, task-oriented organizational structure (see Figure 9.2.2 below).

9.2.2 Requirements and Premises

9.2.2.1 Requirements

The general requirements for quality management result from an interpretation of the requirements specified in one or more standards such as EN ISO 9001, EN ISO 9100 or their equivalent. These are complemented by customer-specific requirements, requirements for technology processes, as well as by direct or indirect rules and regulations under the laws of those countries and communities in which the product is to be used. The customer-specific requirements of major space agencies are summarized in Section 9.2.6.4.

9.2.2.2 Quality Policy and Quality Goals

The quality of products and services forms the basis of the organization's success.

In order to assure and maintain a high level of quality, management usually defines its quality goals as:

- **Optimized customer** satisfaction on a fair partnership basis
- **Failure prevention instead of failure correction (zero-failure goal).**

Quality policy and targets are explicitly defined in the annual business plan in the form of indicators which must be announced and communicated to all management levels.

9.2.2.3 Quality Indicators

Quality indicators are organization- or customer-associated indicators appropriate for evaluating or defining product quality.

Quality indicators can be statistical numbers provided by the organization, project control, quality management, or from production controls. When evaluating such indicators it is essential to take into account the relative complexity and comparability of products. The comparison should therefore only be performed within a specific group of product systems, subsystems, units, or components.

To confirm adherence to the quality targets, the indicators listed in Table 9.2.1 may be used.

Evaluation parameters	Evaluation criteria
Commercial management	
Overall performance	Turnover and other income
Order status	Contract to proposal ratio
Personnel expenditures	Staff versus order status
Average productivity	Productive hours versus overheads
Company expenditures	Rate of investments
Earnings before interest and taxes	Operative result
Cash flow	Income and payment rate
Quality management	
Customer satisfaction	Statistics derived from customer questionnaires
Product failure rates	Deviations or nonconformances, rate of acceptances
Schedule failure rates	Target/actual comparison of milestones or key inspections
Supplier failures	Number of acceptances and refusals, audit rates
Complaint reports	Number, maintenance costs, product improvements/repair
Certifications	Number and type of deviations, certificate validity
Production	
Processability	Process failures versus process certifications
Failure rate	Material and component failures, alerts, rejects

Table 9.2.1: *Evaluation parameters for different areas.*

Additionally, selected data from individual projects can be used:

- Monthly cost and budget control for each project or department
- Monthly booked hours versus hours calculated from work orders
- Comparison of "recurring" to "nonrecurring costs" (as space activities usually involve high, one-off development expenses instead of series products, the ratio of "nonrecurring" to "recurring costs" should be monitored)
- Evaluation of nonconformances, deviations, preventive actions
- Evaluation of deviations from customer requirements.

Benchmarking is a comparison within comparable groups of, for example, organizations, products, processes, methods or services. The following comparisons are often used in projects:

- Customer satisfaction
- Cost
- Schedule and time
- Quality.

Premises are identical and scalable framework conditions such as:

- Exact project description
- Precise project scheduling
- Adherence to defined or required quality standards
- Adherence to budget restrictions.

It is essential that the evaluation and estimation of the indicators mentioned be performed within this framework. The evaluation of commercial indicators is usually related to a business year; the evaluation of quality to a project; and the evaluation of production to a product. Evaluation of these indicators provides useful information about the implementation and effectiveness of the quality management system and how well the organization's quality policy is followed.

9.2.2.4 Customer Satisfaction

Customer satisfaction is a primary goal of an organization. Any deviation from this goal leads to a loss of contracts as it is counterproductive for developing a flourishing business. The most important **criteria** for evaluating customer satisfaction are:

- Rate of delivered proposals to contracted proposals
- Rate of contracts concluded with specific customers
- Rate of product failures and repairs.

Customer satisfaction must be assessed and documented using statistical questionnaires at regular intervals.

9.2.2.5 Manual

The requirements for a quality management system must be translated into the rules and regulations of the organization. In compliance with the ISO standards, all management processes, quality processes and product processes must be described in a dedicated company manual and in associated process procedures. The manual must be regarded as an internal "law" which is binding for both management and all employees. The manual normally describes the following main processes:

1. Management processes
2. Quality management processes
3. Product main processes:
   (a) Marketing and proposal generation
   (b) Predevelopment and development
   (c) Procurement
   (d) Production
   (e) Storage, transport, operation and maintenance.

9.2.2.6 Process Procedures, Work Procedures

An organization's processes are described as written procedures whose updating is the responsibility of managing directors, department heads or team leaders, as appropriate. Process procedures are generically applicable for the generation of all company products. Unique processes which require **product-specific processes** and personnel qualifications are described in specific detailed work procedures. The supervision of all processes including outsourced external processes is under the responsibility of quality management.

9.2.3 The Product Main Processes

Besides the management and quality management processes, specific space product main processes are shown in Figure 9.2.1. In this figure the quality management procedures indicate the correlation of applicable quality control procedures.

9.2.3.1 Marketing and Proposal Submission

All product requirements are described by the customer in a **statement of work** (SOW) and/or in product specifications. Such specifications describe the mission requirements, the technical requirements, the management requirements, the quality requirements and the customer project standards applicable for product realization and verification.

These requirements will be checked following the organization's relevant procedure before incorporation into a product proposal. It is important for product realization within the specified cost and time frame to have a very early declaration of the required **model philosophy**. This has a direct impact on the planning for procurement qualification, material and component costs. Therefore it is mandatory that as part of the proposal phase, all costs and procurement times for the different product models be assessed.

If the procurement is not possible within the project schedule, secondary procurement sources must be considered and alternatives negotiated with the customer.

For software development, either internal or external, adequate qualifications and acceptances must be considered. Special regard must be given to additional costs for software test tools and a test environment.

9.2.3.2 Predevelopment and Development

Any product development is described in a development plan which summarizes the development steps via the different product models, the procurement

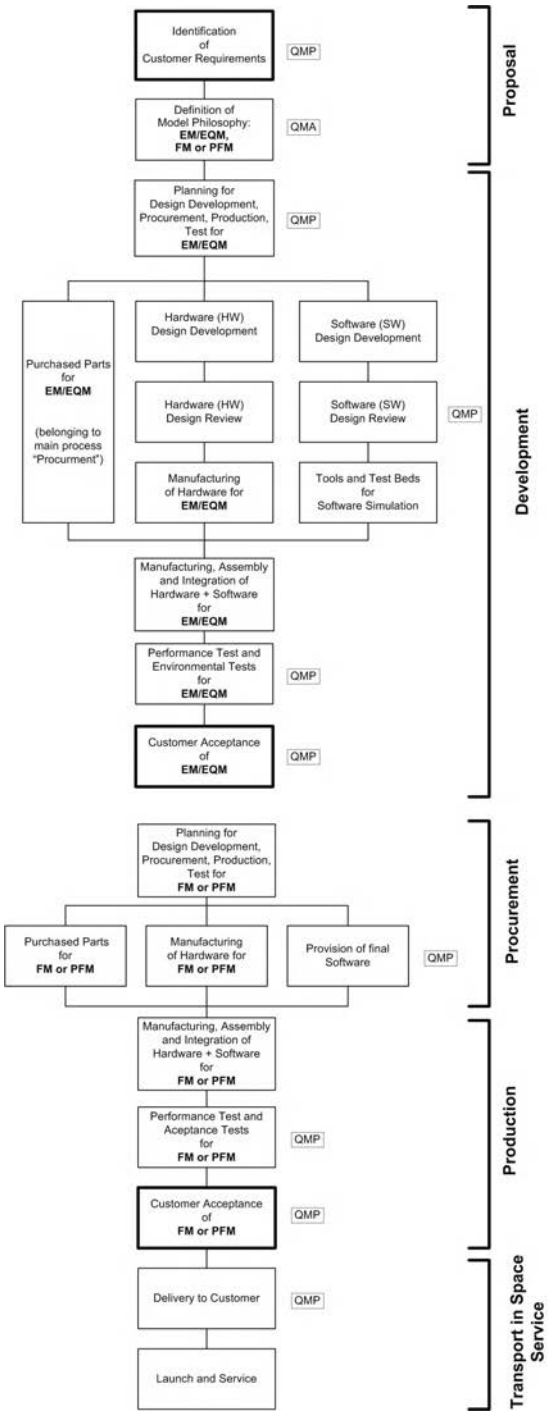


Figure 9.2.1: *The main quality management processes for space activities.*

and production processes, the qualification strategy and the acceptance criteria.

Before development starts, it has to be assured that all customer requirements, for example the design, environmental and operational requirements for the system, subsystems, equipment and units, are completely specified and documented. The product has to be specified by a product tree with configuration items to assure the traceability of the development process.

The development tasks are divided by engineering discipline, as to structure, electrical system, data system, propulsion system and software, and described in detailed work packages.

Product assurance supports and monitors the development according to tasks as described in the product assurance plan. This plan describes the different verification methods for the specified requirements and the applicable models.

9.2.3.3 Procurement

Quality management supports project management in the procurement process, in compliance with the required quality standards and approved manufacturers and suppliers.

Supplier Approval

Manufacturer and supplier are selected in compliance with the product quality requirements. The selection can be either from an existing supplier list provided by the major space agencies, or on the basis of specific technology requirements. For the selection and assignment of suppliers, special concern is given to their ISO certification and their ability to meet customer requirements. In case of any doubts as to their adequacy, or in the case of a new supplier, a **supplier audit** is conducted by quality management. The audit results are reported to the customer on request. Supplier audits must always be completed before the awarding of a contract.

Part of the audit is checking the supplier's compliance with the generic quality requirements of the ISO standards, the product requirements and the supplier's qualification to carry out specified technological processes. Subsequent supplier audits are divided into quality audits, project audits and process audits.

Each audit is directed by quality management, supported by engineering disciplines, if deemed necessary.

Audits are always product associated and must be repeated for each new contract.

The audit performance and results are documented by quality management and included in the organization's list of approved suppliers.

The rating of approved suppliers is based on qualitative and quantitative criteria concerning technical and commercial competence and reliability. The rating must support the selection and may be as follows:

++= best
+= good
-= sufficient
--= not sufficient
?= no experience.

If a supplier is rated as "sufficient" for more than two years, a **control audit** must be performed; if a supplier is rated as "not sufficient" after two years, it must be removed from the supplier list.

ESA maintains a list of **suppliers approved for space products**. They are audited by ESA and can be used by the organization without the need for an additional audit. Any deviation from space-approved suppliers must be reasonable and must in any case be negotiated with the customer.

Quality Agreement

The organization's general quality requirements must be contractually accepted by the supplier in a written quality agreement signed by both parties. The project-specific quality requirements must be contractually accepted by the supplier, who provides a certificate of conformance upon delivery of the product.

9.2.3.4 Production

Supervised production requires production planning and control. **Production planning** describes the production flow and all tasks required for production, such as preparatory work, work scheduling, integration and test planning. Production planning is performed by the production manager supported by system engineering and product assurance staff.

To assure the quality of the production process, the following preconditions must be fulfilled:

- Planning and definition of the fabrication processes
- Maintenance of production facilities and manufacturing tools
- Adequate work environment
- Adequate process procedures
- Personnel qualification and training.

Each single process, for example machining, nondestructive inspection or functional testing, is described in a fabrication instruction or test procedure with a declaration of tools, materials and personnel. Each process execution is documented in a report to be checked by those responsible for product assurance.

The production manager **checks the manufacturing** on site according to customer, technology or quality-specific criteria. The selection of manufacturers and test facilities is a combined activity involving engineering disciplines, product assurance and project management.

All steps and processes needed for production, integration and testing are identified and scheduled in a manufacturing, integration and test plan. Key characteristics, inspections and acceptances are identified and scheduled in a dedicated inspection plan. For space products, environmental restrictions and contamination controls must be respected.

9.2.3.5 Storage, Transport, Operation and Maintenance

Storage and transport of space products must be in accordance with special procedures under environmental control. To avoid inappropriate handling, all mechanisms, interfaces and control elements have to be clearly marked and identified according to project requirements or international technical rules and industrial safety standards. Transportation and handling must be described in specific manuals with identification of warnings of safety hazards, special precautions and dangerous goods by references to public authorities. Special customer requirements must be identified as part of the product specification.

For delivery, the regulations of the user organization, country or launch site must be respected. A

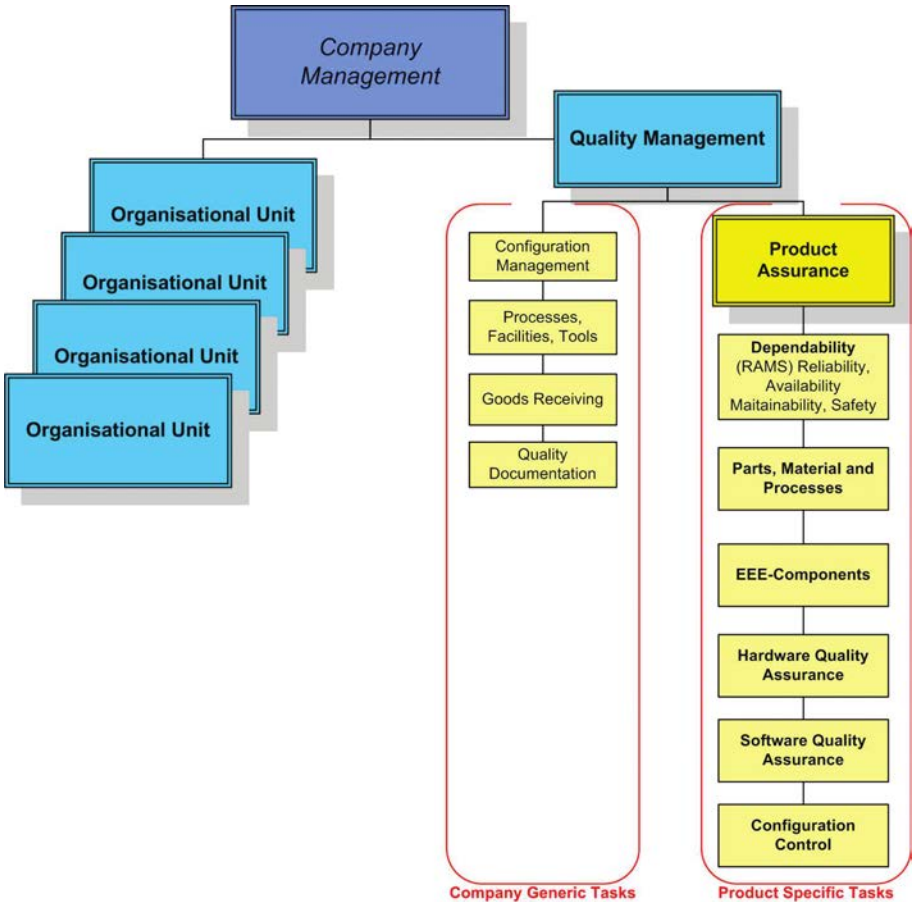


Figure 9.2.2: *A typical quality management structure for a space organization.*

space product in orbit has to be maintained from the ground by the control center. This assures operational availability through regular monitoring of specific characteristics such as data transfer, orbit maneuvers, or housekeeping data.

The execution and results of such maintenance must be documented in service manuals and historical records. Whenever a change of configuration is required, especially for operational software, procedures for change management and configuration control apply.

9.2.4 The Organization of Quality Management

The central quality management is placed under the organization's overall management and reports directly to the chief executive officer (CEO). The product assurance function can be either centrally organized under the quality management, or decentralized and allocated to engineering functions. It is possible that, for example, hardware quality assurance is allocated to the "production" process and software quality assurance to the "development" process. The advantages of centralized product assurance are concentration of the quality knowledge base, coherent guidance of all product assurance tasks for all products, and the compilation of all quality data in one quality documentation center.

Irrespective of these organizational options, reporting to the central quality management and the CEO is mandatory.

A practical example of organizing quality management and product assurance is depicted in Figure 9.2.2.

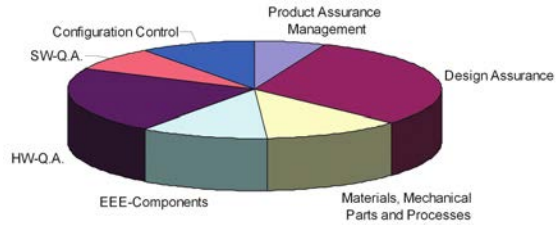
9.2.5 Product Assurance (PA)

Each project is identified by a unique statement of tasks, cost and schedule limitations. Subsequently the identification of individual and specific quality measures must be planned, witnessed and documented by product assurance. The product assurance tasks are derived from generic product quality assurance measures and customer-specific quality requirements.

The main interdisciplinary tasks of product assurance are:

- Coordination of PA personnel, tasks, **cost and schedule** for all product phases
- Coordination of resources, personnel training and certification
- Maintenance of quality **documentation**, standards, standard practices and customer requirement specifications
- Preparation of project **product assurance plans** (tasks, personnel, cost, schedule, outputs)
- Witnessing of internal and external **product and project audits**
- **Support** of engineering disciplines and project management
- Focal point of contact for all **customer questions** regarding quality aspects
- Coordination of procurement plans, export licenses and alerts
- Support to project management, configuration and documentation management (CDM)
- Nonconformance control and change control.

Realization of these tasks requires an organizational structure which represents approximately 5 to 10% of the staff members. This number is necessary for fulfilling all the quality obligations for all products and projects of the organization. For the support of all product main processes, the apportionment of product assurance tasks can be as follows (see Figure 9.2.3).



Product assurance coordination	6%
Design assurance	31%
Materials, mechanical parts and processes	12%
EEE components	11%
Hardware quality assurance	22%
Software quality assurance	8%
Configuration control	10%.

Figure 9.2.3: *Average distribution of product assurance tasks.*

9.2.5.1 Design Assurance

Design assurance describes the analytical **evaluation** and auditing methods leading to the **attestation** of corrective measures for the designated product characteristics. The identification of technical risks (critical items; see Figure 9.2.4) which could lead to a restriction of the reliability, availability, maintainability or safety of the system is described as **dependability**.

Dependability covers all reliability, availability, maintainability and safety analyses to assure that:

- All technical risks are identified and evaluated quantitatively and qualitatively.
- All risk consequences are evaluated.
- Reductions and controls are defined and implemented.
- Residual risks are monitored by the risk management process.
- All checking and verification measures comply with the reliability, availability, maintainability and safety requirements.

Design assurance starts with the conceptual product phase and continues in the predevelopment and development phases up until the final design. The analyses start on the product system level and are performed "top-down" to the component level. The implementation of corrective measures follows the reverse approach, "bottom-up," to assure consolidation of requirements from the lowest level to the system level.

9.2.5.2 Reliability

Reliability assurance activities comprise the qualitative **identification** and analysis of failures as well as the quantitative prediction of the **probability of failure occurrence**. The basis for these analyses is a detailed

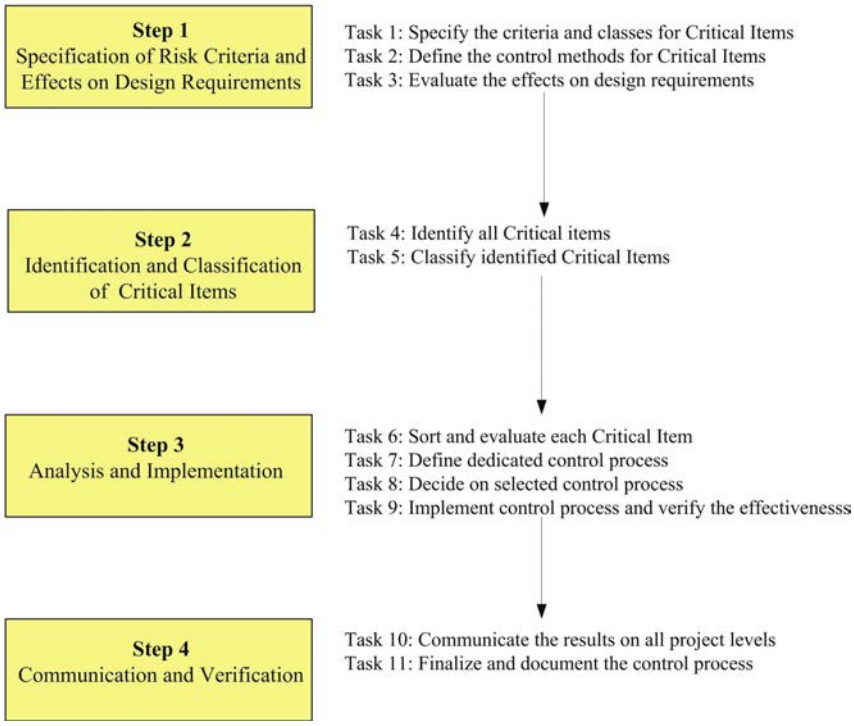


Figure 9.2.4: *Control of critical items.*

functional analysis for each element of the product system configuration in the form of an FMECA (Failure Mode Effects and Criticality Analysis) supported by the results of a numerical reliability analysis. As a result of such a risk evaluation process, redundancy concepts and design checks are defined, especially for system-inherent "single points of failure," as well as concepts for preventive and corrective maintenance.

Reliability-critical elements (critical items) and critical items from other analyses, for example safety analysis, fault tree analysis, operations and structure analysis, must be considered as part of reliability assurance.

The main tasks of reliability assurance are:

- Performance of analyses during product Phases A, B, C/D
- Analysis of failure scenario apportionment, predictions, analysis and availability
- Failure mode effects and criticality analyses (product and process FMECA)
- Fault tree analysis (top-down or bottom-up)
- Parts count and parts stress analyses for electronic components
- Duty cycle and worst case analyses
- Hardware/software interaction analyses (HSIA)
- Definition of actions and recommendations for detailed risk assessment, risk elimination or risk reduction
- Identification, monitoring and verification of critical elements (critical items)
- Evaluation of redundancy concepts and single points of failure.

9.2.5.3 Availability

The activities of availability assurance comprise analyses to determine the **availability of a space segment and a ground segment**. The major part of the availability analysis is verification of the operational availability of the complete system for specified times under predicted conditions with regard to planned dormancy, maintenance and repair times.

The availabilities of the space segment and the ground segment naturally differ. The availability of the space segment is mainly dependent on the contact times; the availability of the ground segment

is dependent on maintenance and repair. Maintenance can be either preventive (which means planned) or corrective (which means unplanned). The basis for the availability analyses is the calculated reliability numbers of both segments for the predicted operational life, influenced by time for maintenance, repair and administrative decisions, as well as time for the logistics related to hardware and software.

9.2.5.4 Maintainability

The main activities of this discipline comprise the evaluation of **maintainability concepts** leading to maintenance procedures for human and unmanned space products and ground segments.

This includes the preparation of maintenance analyses under the aspects of human safety and system reliability. Here the role of product assurance is limited to a supporting function focused on the control of potentially hazardous maintenance functions and operations (manual or automatic) and the approval and release of maintenance procedures and tools.

The main tasks of maintainability assurance are:

- Identification of limited-life items
- Evaluation of accessibilities for maintenance (only for human space products)
- Development of repair and maintenance concepts
- Evaluation of availability of spare parts, procurement and logistics planning
- Performance of duty cycle and operation cycle analyses.

9.2.5.5 Safety

The main activities of safety assurance comprise the definition of safety requirements for a space product, the preparation of safety analyses, the implementation of safety measures and design controls in the development phase, the verification of safety operations, and the performance of customer-required safety reviews. The accomplishment of a safety program is based mainly on the requirements specified by the space agencies and the launch authorities. The safety activities are summarized in a reviewed **safety document** on the different hazard potentials in human and unmanned space products and associated ground segments.

The main tasks of safety assurance are:

- Safety reviews through project Phases A, B, C/D
- Identification of safety requirements for personnel, material and operations
- Safety discussions and hazard reports according to NASA NSTS 1700.7 B + ISS Addendum
- Preparation of safety data packages and single safety analyses
  - Preliminary hazard analyses (PHA)
  - Subsystem/system hazard analyses (SHA)
  - Operating hazard analyses (OHA)
  - Warning time analyses (WTA)
  - Residual hazard lists
- Implementation of safety programs according to NSTS/ISS 13830 and JSC 11123
- Implementation of safety programs according to customer programs
- Implementation of control in design and operations
- Definition of controls for residual hazards.

The functional and operational safety analyses serve to monitor safety-critical functions and operations, identify failure propagation and determine safety criticality factors and associated hazard controls. Besides product-associated safety requirements, the national safety rules and regulations for industrial and human safety must be respected.

9.2.5.6 Parts, Materials and Processes

The main activity in this discipline is the specification of appropriate **quality standards** for parts, materials and processes to be used for a space product with respect to their predicted operational life, operational environmental conditions and hazard potential.

Selection, quality standard checks and acceptance of metallic and nonmetallic materials and the associated production and inspection procedures must be verified for each construction element before use in a space product. The material properties, the production process and the material-inherent potential for hazards must be evaluated to define detailed acceptance processes and checks, for example for corrosion, flammability, out/offgassing and toxicity. The main monitoring tasks are:

- Performance of qualification and test programs for parts and materials

- Identification and scrutiny of new technologies and critical manufacturing processes
- Qualification of processes and personnel, process audits, process acceptances
- Maintenance of material databases, standard practices, standards and alerts
- Preparation of materials list, mechanical parts list, process lists
- Preparation of requests for approval (RFA) of materials
- Performance of flammability and off/outgassing tests
- Review of fracture-critical items and their verification
- Review of critical (dye-penetrant, X-ray, ultrasonic) processes for non-destructive inspection (NDI)
- Review/support for environmental tests (e.g., thermal vacuum tests, thermal cycle tests).

Offgassing

All nonmetallic materials release trace contaminants into the surrounding environment; the extent to which this occurs depends on the nature of the material concerned.

In the closed environment of humans on-board a spacecraft, contaminants in the atmosphere are potentially dangerous because of their toxicity, and close monitoring is therefore required.

Offgassing is the evolution of gaseous products in an assembled article when it is subjected to slight radiant heat in a specified test atmosphere.

An offgassed product is an organic or inorganic compound evolved from a material, assembled article, experiment, rack, etc.

Molecular Outgassing

Space system materials outgas in the vacuum of space; generally the level of space vacuum does not affect the outgassing below pressures of 10^{-2} mbar.

The materials used for space systems are normally selected based upon low outgassing criteria:

A: For normal applications the outgassing criteria are based upon the Micro-VCM test (ND-9 or ECSS-Q-70-04) and amount to recovered mass loss (RML) < 1% and collected volatile condensable material (CVCM) < 0.1%.

B: For optical instruments it is generally advisable to have more stringent outgassing requirements and to perform baking tests on the relevant hardware. Materials such as cadmium and zinc have high vapor pressures and may deposit metallic films on adjacent surfaces.

C: The use of these volatile metals is basically forbidden, especially for temperatures above room temperature.

9.2.5.7 Electrical, Electronic and Electromechanical (EEE) Components

The main activities in this discipline are the specification of **quality standards** for EEE components to be used for a space product, considering its predicted operational life, operational environmental conditions and hazard potential. Special concern is given to the procurement planning, which depends on the export restrictions defined by governmental authorities for national and international component suppliers.

The main monitoring tasks are:

- Definition of component quality standards according to international space standards
- Compatibility check of the selected standards versus product applications
- Procurement planning with regard to the supplier sources selected by the space agency
- Identification of export restrictions, procurement restrictions and long-lead items
- Evaluation of requirements concerning radiation tolerance and hardness
- Evaluation of derating limits according to space agency and component-specific limits
- Evaluation of EMC classes for cabling and connectors in their system application
- Evaluation of bonding, grounding and shielding concepts
- Evaluation of printed circuit board design according to space agency requirements
- Processing of parts approval documents (PAD) and requests for approval (RFA)
- Preparation of declared components lists
- Component alert check.

In cases where space-qualified components are not available for certain applications, the use of **commercial off-the-shelf** (COTS) components may be

necessary. For their approval additional and dedicated delta qualifications are mandatory (Figure 9.2.5). The qualification process must in any case be released by the customer and adhere to the sequence below.

9.2.5.8 Hardware Quality Assurance

The activities for hardware quality assurance include approval for the implementation audit of customer quality requirements throughout the complete product life cycle from design development, production and qualification until final acceptance and delivery.

Design monitoring, production inspections, qualification witnessing and acceptance procedures are specified in space agencies' standards (e.g., ESA Standard ECSS-Q-20, *Quality Assurance*).

The main activities are:

- Traceability of the product life cycle from design development to acceptance
- Auditing of subcontractors, manufacturers and suppliers, procurement review
- Industrial safety monitoring of facilities (e.g., clean-room conditions according to international ISO standards)

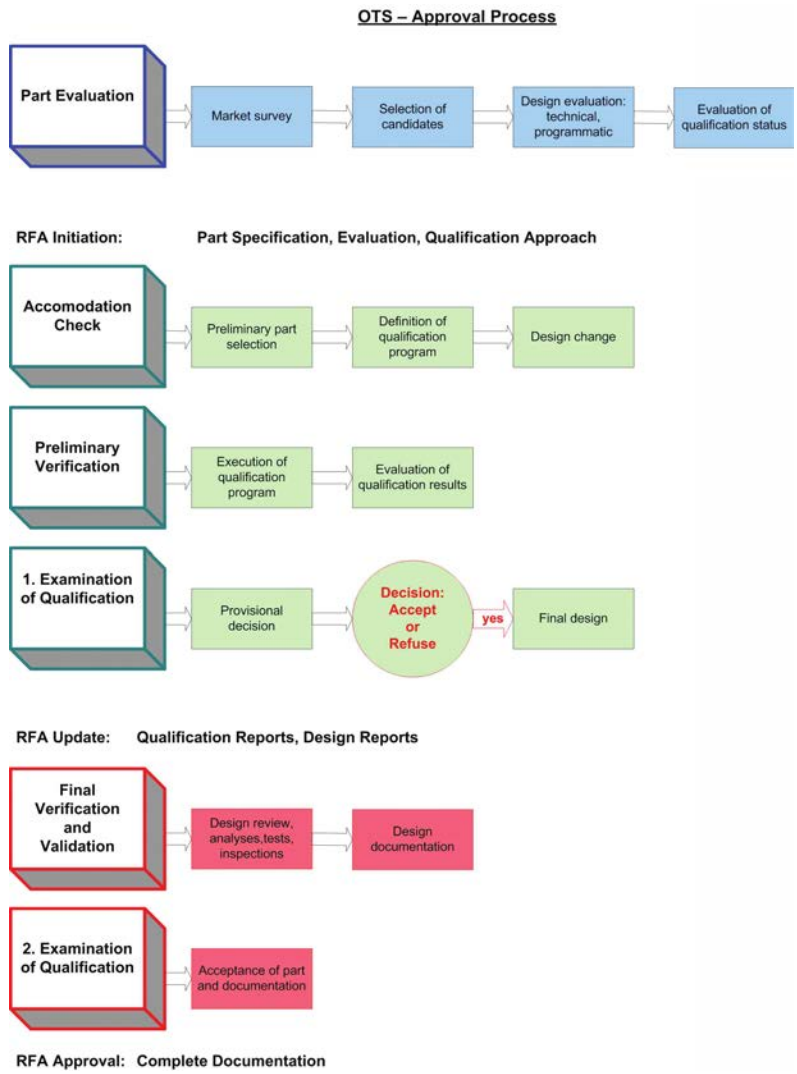


Figure 9.2.5: *Selection process for commercial off-the-shelf components.*

- Metrology of measurement and test equipment via databases
- Review of manufacturing, operations, handling, storage, cleanliness and transport procedures
- Witnessing of hardware tests, acceptance tests for space products
- Performance of inspections (e.g., mandatory/key inspection points on all product levels)
- Review and approval of test plan, test procedures, test reports
- Confirmation of configuration status "as-designed" versus "as-built"
- Verification check of customer requirements
- Standard inspection for hardware receipt, handling, storage, checkout, packaging and shipment
- Metrology check and monitoring, review of AIT facilities, cleanliness levels, process qualifications,
- Preparation of inspection reports/records, as-built certifications, cleanliness certificates, documentation of hardware life cycles (e.g., log books)
- Support for preparation of end item data packages for the different models.

9.2.5.9 Software Quality Assurance

Software quality assurance activities, as for hardware quality assurance, also comprise the implementation review of customer quality requirements throughout the complete product life cycle from design development, production and qualification until final acceptance and delivery, but as a common approach with software engineering (see Figure 9.2.6). Development, qualification, implementation and validation of software are preferably performed according to a **project associated modeling (V-Model)** or a **generic phase model** according to ESA Standard ECSS-Q-80B, *Software Product Assurance.*

All phases must be described in a company software quality manual and more specifically in the project product assurance plan for the following phases:

Phase 1: Software requirements specification
Phase 2: Software functional analysis (architectural design)
Phase 3: Software design (detailed design)
Phase 4: Software implementation (simulation)
Phase 5: Software integration (on hardware)
Phase 6: Software testing
Phase 7: Software acceptance
Phase 8: Software maintenance.

The main quality assurance activities for software are:

- Traceability of the product life cycle from design development until acceptance
- Confirmation of implemented software engineering standards
- Performance of inspection (e.g., mandatory/key inspections on all product levels)
- Witnessing of software acceptance testing
- Approval of software implementation
- Comparison of design development versus production status (as-built versus as-designed)
- Verification check of customer requirements
- Preparation of acceptance data packages and collection of product quality documentation
- Maintenance of the software quality manual and procedures
- Review and approval of an independent software verification and validation (ISVV) plan
- Analysis of software safety as part of the system safety analysis, depending on the application
- Software version control by approval of software release orders
- Software anomaly and non-conformance control.

9.2.6 Product Assurance for a Project

The product assurance tasks for a project derive directly from the customer requirements in the statement of work (SOW) and associated specifications. These define the technical product requirements, the model philosophy, the deliverable items, as well as the product documentation. All product assurance tasks are related to the project processes and procedures and documented in the product assurance plan.

9.2.6.1 Project Phases

The project processes are nearly identical to the generic company main processes. Depending on the customer contract, they are more detailed and represent the main project phases. These phases represent in parallel the work breakdown structure (e.g., for

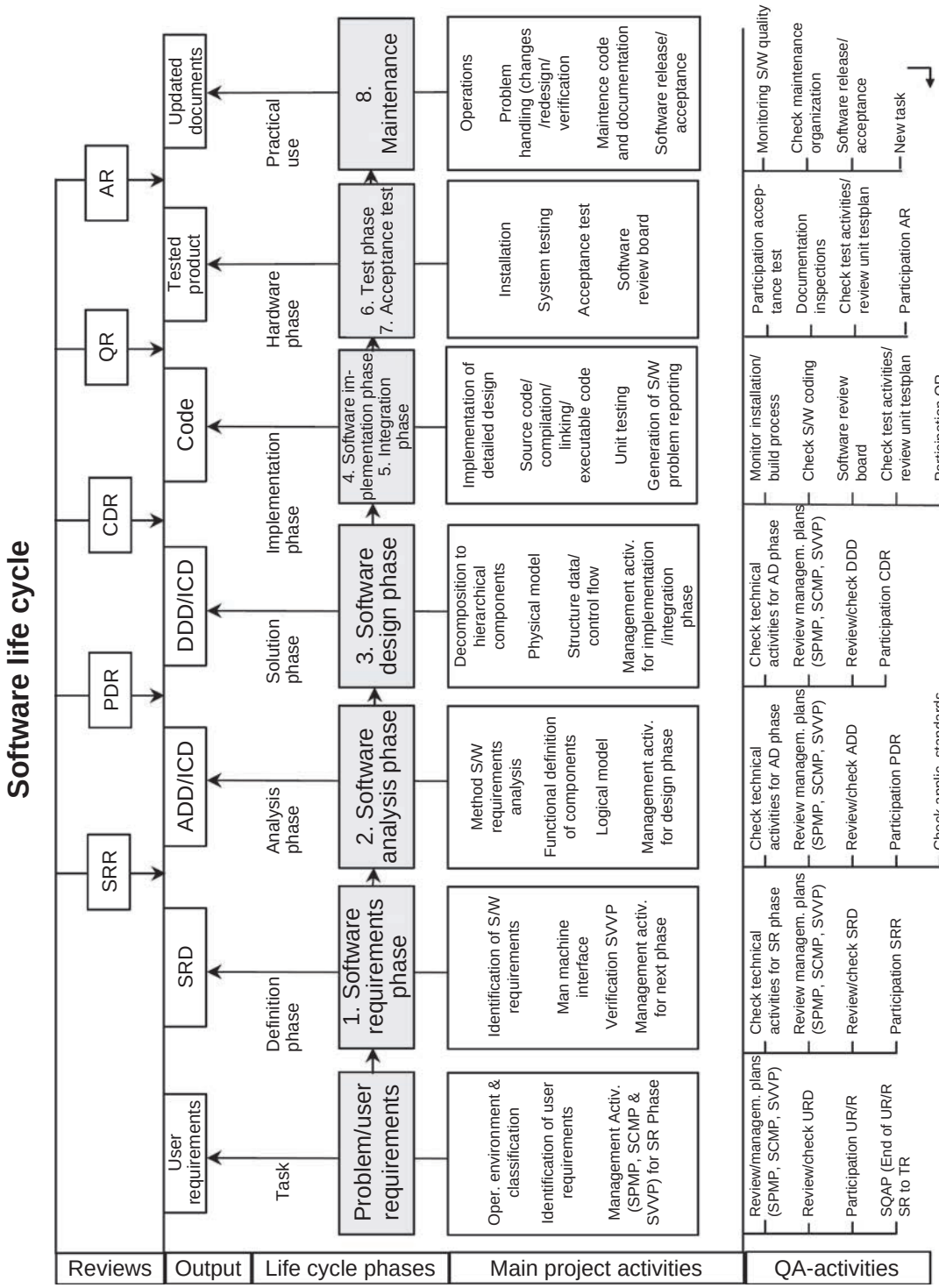


Figure 9.2.6: *Software life cycle.*

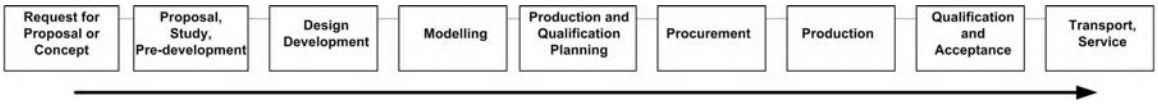


Figure 9.2.7: *Typical course of project phases.*

engineering, procurement, integration and testing) and allow allocation to lower level project element and work package descriptions.

Depending on the project type, all or only some phases are realized. It is the responsibility of project management to plan these phases (Figure 9.2.7) and provide detailed project plans.

9.2.6.2 Project Interfaces

Effective product assurance relies on work contributions from other operational functions in the organization and from the project. The technical and administrative interfaces are described in the **work breakdown structure** (WBS), their associated **work package descriptions** (WPDs) and the common project milestone reviews and documentation.
The main project interfaces are:

- Project management
- Configuration and data management
- Engineering disciplines
- Procurement and manufacturing
- Assembly, integration and test
- Cost controlling.

9.2.6.3 Product Tree

The product tree describes the top-down configuration of the product on the system, subsystem, equipment and component levels, each identified by a dedicated configuration item (CI) number (Figure 9.2.8).

9.2.6.4 Requirements and Standards

As already mentioned, the quality requirements for a product are issued by the customer, preferably by the relevant space agency. The following associated requirements must be respected:

- Requirements from the launch authority and launcher interfaces
- Scientific or commercial operational mission requirements
- National laws and regulations concerning industrial safety, hazardous material control, export restrictions
- The organization's quality requirements and procedures.

ESA

The quality and safety requirements for human and unmanned space products are divided into three main groups according to ECSS standards:

- ECSS-M-00 *Management for Space Products*
- ECSS-E-00 *Engineering for Space Products*
- ECSS-Q-00 *Product Assurance for Space Products.*

NASA

The quality and safety requirements differ for human and unmanned space products and are associated with the particular launch vehicle and responsible NASA centers:

- JSC (Lyndon B.) Johnson Space Center, Houston, Texas
- MSFC (George C.) Marshall Space Flight Center, Huntsville, Alabama
- KSC Kennedy Space Center, Orlando, Florida.

The requirements are often established and decentralized, and are composed of selections from the following documents:

- NASA-STD *NASA Standards*
- NSTS *National Space Transportation System (NASA)*
- NHB *NASA Handbook (Quality and Safety)*
- SSP *Space Station Program (Quality and Safety).*

DLR

The quality and safety requirements for human and unmanned space products are included in two main volumes:

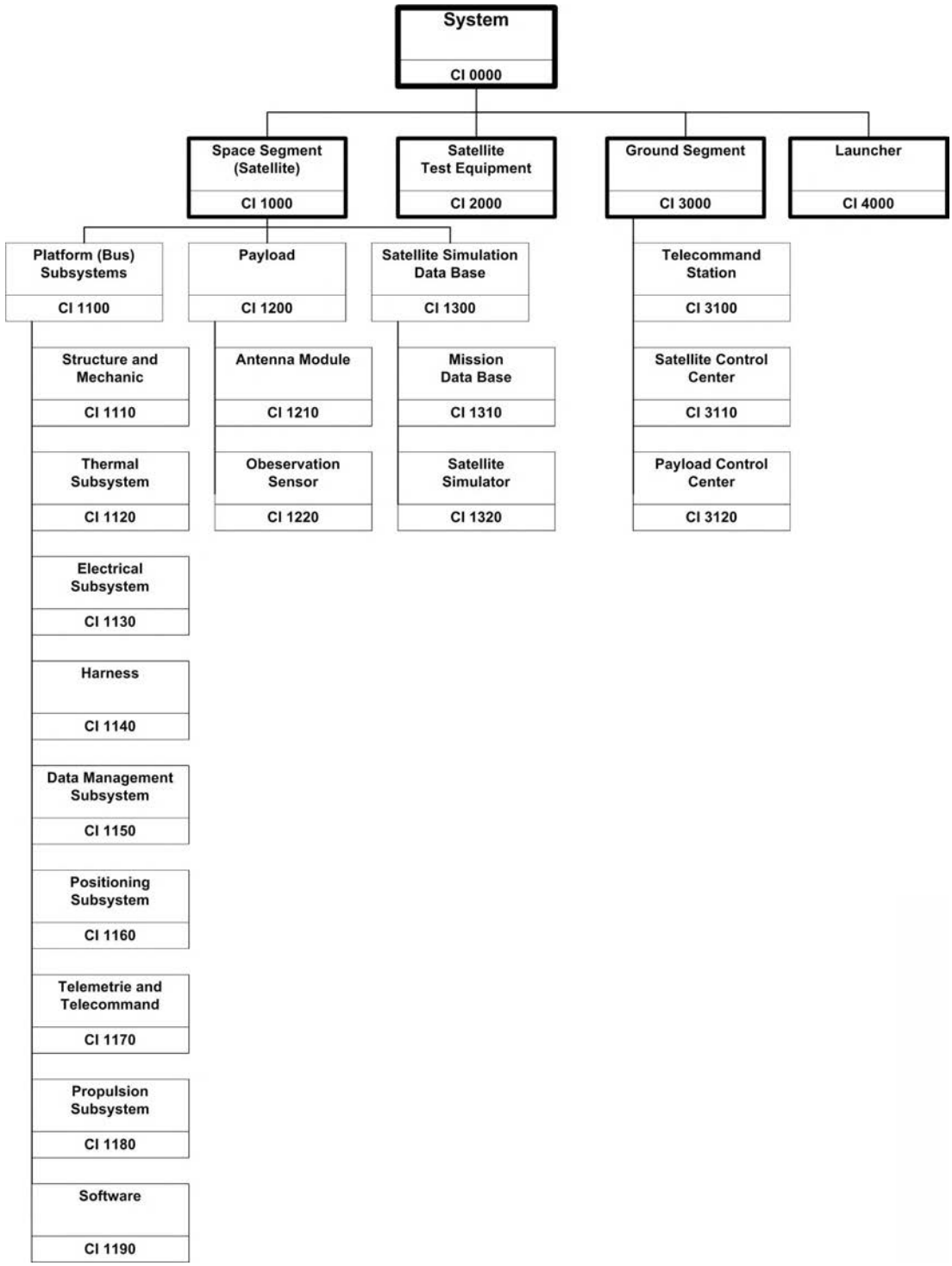


Figure 9.2.8: *Typical product tree for a satellite, showing the configuration items.*

- DLR-RF-PS-001 *Product Assurance and Safety Requirements for DLR Space Projects (Tailoring Catalog)*
- DLR-RF-PS-002 *Product Assurance and Safety Requirements for DLR Space Projects (Document Requirement Definition, DRD)*

9.2.6.5 Customer Requirement Specification

Usually the product requirements are specified in the customer requirement specification included in the statement of work (SOW). These are accommodated by the contractor in the form of a design specification for the system, the subsystems, the operations and the associated work package descriptions (WPD). It is essential that all requirements be completely identified, collected and documented before the contract. They are contractually committing and must be explicitly verified until the project comes to an end.

The **specifications** contain the design requirements, operational requirements, environmental requirements, quality standards for parts, materials and components, as well as the deliverable product models. The requirements must be broken down to all product levels, as system, subsystem, equipment and component levels. Special concerns have to be addressed concerning the functional and physical interfaces between these levels, the payload and the interface in human space projects. Incomplete, unclear or contradicting requirements must be negotiated with the customer before the project starts.

9.2.6.6 Selection of Quality Standards

The selection of appropriate quality standards is primarily oriented along the functional requirements, the environmental loads and the predicted operational lifetime. The standards may differ, depending on the maintenance concepts, especially for human space products. The quality standards differ for the required product models, of which the flight model has the highest quality standards.

Additional selection criteria derive from the reliability and availability requirements and the chosen redundancy concept. In the case of sufficient redundancies the quality standards for components may be reduced, but on the other hand they may have to be increased in the case of insufficient redundancies. All selected quality standards must be checked by product assurance and documented in compliance lists.

9.2.6.7 Product Model Philosophy

The definition of the model philosophy is especially important for product development and production. It has a direct impact on qualification planning, procurement planning and hardware cost. The selected quality standards must be related to the different product models. Model hardware cost and hardware procurement sources as well as the procurement time have to be identified early in the project proposal phase. In order to meet the project schedule it is in any case recommended that alternative hardware procurement sources be identified.

Engineering (Qualification) Model and Flight Model

For this "two-model" philosophy the **engineering model** (EM) serves as a development model which is identical to the **flight model** (FM) in form, fit and function. In exceptional and reasonable cases it is permissible to use lower quality standards for EEE components. Therefore the failure risk is relatively high but controllable. The EM is used for flight qualification testing (EQM) including environmental testing. Design improvements, modifications and repairs are allowed.

The subsequent FM is a reproduction of the qualified EM but with flight-standard hardware. The FM undergoes acceptance testing with lower testing levels than for the EQM. As depicted in the schematic of Figure 9.2.9, the risk of failures in the FM is significantly reduced by this "two-model" philosophy. Residual but controllable risks can occur during system integration when the interfaces between the different subsystems and payloads are verified. These risks cannot be significantly reduced and remain even if additional models are used.

Prototype Flight Model

Typical of the **prototype flight model** (PFM) is the application of flight-standard hardware without exception and without any upstream qualification

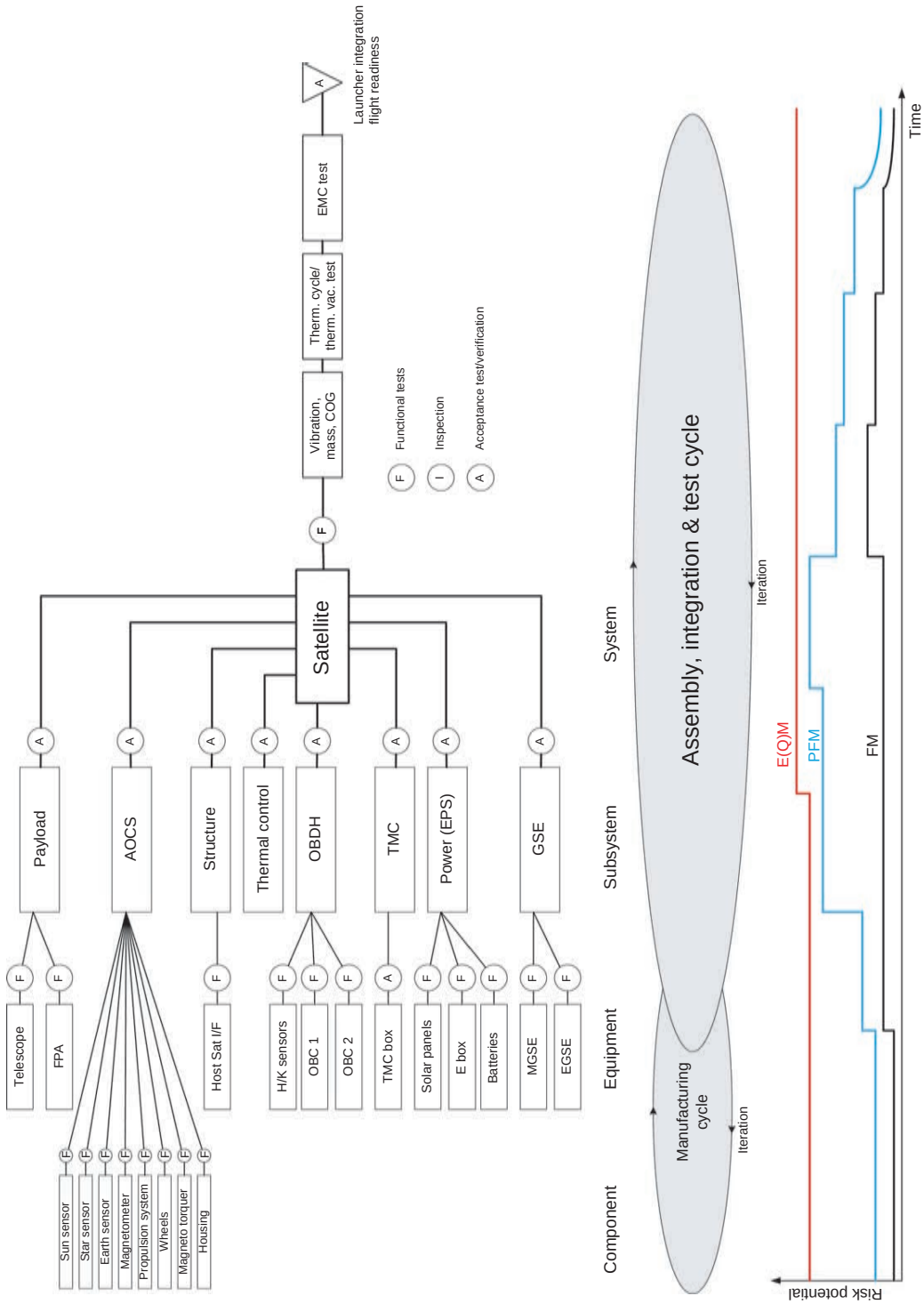


Figure 9.2.9: *Abstract integration flow for a satellite with associated risk potential dependent on the selected model philosophy.*

model. The failure risk on the element and component levels is lower than in the EM, presuming that the environmental qualifications have been performed on the component, equipment and subsystem levels before PFM integration. The main risk now occurs during integration of these elements for functional interaction. Experience from an upstream EQM does not exist, which increases the failure risk until functional testing on the integrated system level. The risk during environmental tests with a combined qualification/acceptance level is controllable because of the previous qualification-level tests on the component, equipment and subsystem levels.

Failure probability and total risk are explicitly higher than the risk associated with the "two-model" philosophy.

9.2.6.8 Project Milestones

Monitoring the product life cycle is defined contractually and is performed by specified project milestone reviews. The complexity and purpose are specified by the customer and must be reflected in the project plan, the product assurance plan and the project schedule.

The milestone reviews (Table 9.2.2) have to monitor progress of the project, completion of the schedule and the status of the requirements verification in order to clear the way for the subsequent product cycle.

9.2.6.9 Product Assurance Planning

The planning of the product assurance task must ensure the early identification of the product assurance work packages in order to guarantee product realization. The planning has to be performed as early as possible, preferably during the proposal or study phase in order to identify the required extent of product assurance methods, personnel, personnel qualifications, resources and costs. The product assurance activities are described in detailed work package descriptions (WPDs), as for all other project activities according to the work breakdown structure (WBS) associated with the product life cycle.

The planning is performed in two steps: the timely association of all product assurance activities with the product life cycle in a product assurance task flow; and the technical and programmatic activity description in a written product assurance plan. A precondition is knowing the customer's quality requirements, project schedule, model and qualification strategy, and deliverable quality documentation.

The identification and association of product assurance activities and costs are simplified by separation into:

Hardware costs:	Costs for product elements and procurement, support equipment, tools
Software costs:	Licenses, simulators, tools
Labor costs:	Nonrecurring and recurring,
Qualification costs:	Internal and external test facilities, spares
Task methodology:	Computer, documentation, presentations.

The single product assurance expenditures can now be easily allocated to the work packages. The total

Abbreviation	Name	Description
RR	Requirements Review	Completion of requirements selection according to the specifications
PDR	Preliminary Design Review	End of initial design phase, verification of functionalities
CDR	Critical Design Review	End of detailed design phase, clearance for production
QR	Qualification Review	End of performance and environmental qualification phase
AR	Acceptance Review	Product (hardware and software) and documentation acceptance
FRR	Flight Readiness Review	Release for launcher integration
LRR	Launch Readiness Review	Release for launch

Table 9.2.2: *Project milestones.*

expenditure for all product assurance activities is usually 4–10% of the total product cost, depending on the product type (study, human/unmanned space product), development risk, quality standards, qualification strategy, model philosophy and vertical range of manufacture.

9.2.6.10 Preconditions for Planning

For the planning of activities a distinct product description and identification of product elements by the product tree for system, subsystem, equipment and components is essential.

The following information is required:

- Management requirements, organizational structure for contractor and subcontractors
- Customer requirements, hardware and software standards
- Mission goal and operational lifetime
- Model philosophy for space segment, ground segment and ground support equipment
- Environmental conditions, operational conditions
- Launcher
- Project schedule
- Overall cost
- Export restrictions
- Deliverable documentation.

9.2.6.11 Product Assurance Task Flow

The **product assurance task flow** (see Figure 9.2.10) is a supporting tool for preventive planning. It simplifies the timely association of PA activities to project activities in a schematic form. The allocation of PA activities is performed for each phase of the product life cycle, demonstrated here by using the example of "design development" (in the tinted boxes).

Product Phases/Processes: All phases of the product life cycle (according to DIN 55350) in a timely sequence from the proposal to delivery of the product.

Elements: All work steps of the selected product phase and work activities, as described in the work breakdown structure (WBS).

Project Activities: All project activities of the selected product phase as described in the work package descriptions (WPDs).

Product Assurance Activities: All PA activities and methods associated with project activities (see tinted boxes).

9.2.6.12 Product Assurance Plan

The PA plan reflects the understanding of customer requirements and referenced documents and describes the associated product assurance program throughout all project phases.

The PA program describes the technical and quality activities to check the product life cycle for design development, manufacturing, qualification, integration and acceptance.

The plan is applicable to the contractor with respect to the customer, and is part of the subcontracts with the industrial partners for the space product. The inherent quality policy is also used as a guideline for selected manufacturers and suppliers of equipment, units and components.

The plan's sections are considered to be requirements and must be selected and implemented as appropriate in the associated equipment specifications.

The PA program normally starts with Phase B for the product phase and continues in Phase C/D. Specific quality assurance tasks may be continued in Phase E, the operational phase.

The PA tasks correlate to the project milestone plan and schedule, for example design reviews, progress meetings, safety reviews and acceptance reviews.

The PA program may be supported by the organization's in-house regulations and procedures according to ISO 9100 as implemented by the prime contractor and the subcontractors.

The subcontractors must declare compliance with the system's PA plan by means of a subcontractor product assurance compliance matrix supported by a sub-PA plan based on their in-house rules, regulations and procedures, both to be approved by the prime contractor before the project starts.

The PA plan becomes a contractually binding document, which includes the following:

- Applicability and directions
- References to customer requirements and frame conditions
- PA organization and personnel associations

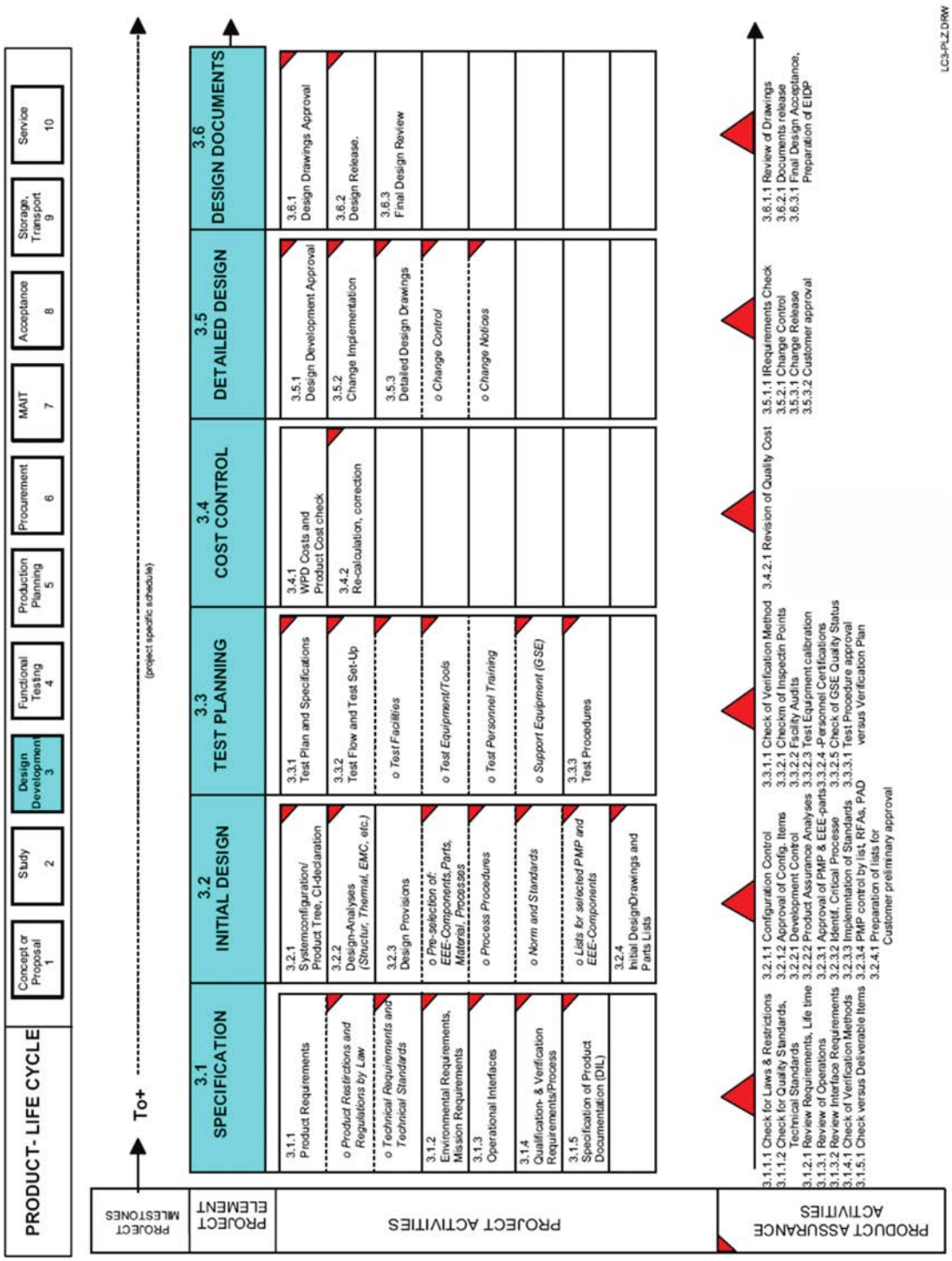


Figure 9.2.10: *Product assurance task flow (example of design development).*

- Description of PA disciplines and activities
- Interfaces to project disciplines
- Description of subcontractor, supplier and manufacturer control
- Reporting duty, reporting content and criteria for milestone completion
- Procurement planning
- Failure, deviation and change management
- Configuration check
- Verification check
- Content of the quality documentation deliverable.

9.2.7 Risk Management

Risk management is an iterative process that is applied to company processes as well as to products, according to space agency standards. The risk policy for the company assures that the risk inventory and estimation will be performed for business processes and survival risks on a regular basis.

Business risks must be identified by forewarning indicators and thresholds for the organization's main processes. Identification is supported by the use of the quality indicators.

Risk management assures the implementation of control processes and corrective action in order to avoid or reduce risk. A risk is any event that can have a negative impact on the business or the product, with a certain probability of occurrence, as well as any future event on which the impact is not presently known.

The risk management process must be described in a separate risk management plan which essentially consists of the following activities.

Risk Identification: The first step in risk management is to identify all individual risks that can affect the project's objectives in terms of cost, schedule and technical performance. The definition of the risk scenario presupposes the assumption of all possible events impacting the product negatively (i.e., the nonachievement of product objectives) and the associated potential causes of those risks.

Risk Assessment: The purpose of this step is to determine the magnitude of the individual risks and to rank them with respect to cost, schedule and technical performance. The magnitude of risk is calculated as the product of a probability score and a severity score.

A risk mitigation table details the criteria used to evaluate the probability of occurrence and the consequent severity of an event identified as a risk item for the program.

Risk Reduction: The ultimate purpose of risk management is to reduce the magnitude of risk by proper mitigating actions in order to achieve the product's objectives; it involves revising the project schedule, budget, scope or quality.

Risk Monitoring: The purpose of this step is to observe the monitoring process and corrective actions.

Communication and Acceptance: The purpose of this step is to track, monitor, update, iterate and communicate the risks in order to manage them.

9.2.8 Configuration Management

Configuration management (CM) is the progress-oriented management and audit of requirement and design documentation data files, and of the items produced.

CM processes are applied for:

- Hardware product configuration
- Software product configuration
- Documentation.

The main objectives of CM are coordination and synchronization of distributed development and maintaining control over the product life cycle and the maintenance of product consistency despite various planned or unplanned changes. CM provides a methodical approach by applying methods for:

- A well-structured product breakdown and reference configurations
- An unequivocal definition of design-to, produce-to, test-to and operate-to documentation and products
- Effective and efficient change management for active control
- Status accounting and reporting for progress judgment and monitoring
- The minimization of administrative activities by centralized approaches
- An assurance that everyone working on the project (including subcontractors) at any time is using the

correct and accurate documentation and products (e.g., computer programs).

By the above means, CM provides the basis for a coordinated and reproducible process and makes an essential contribution to the PA program.

Documentation management provides the capabilities and regulations for establishing and handling all project documentation in order to assure authorized preparation, update, storage and delivery based on CM-controlled procedures, independent of the medium on which the information is recorded.

CM must be described in a product-dedicated configuration management plan.

9.2.9 Failure and Change Management

Failure management monitors all deviations from the contractual product baseline. Each identified non-conformance, for instance manufacturing failures, must be reported in a nonconformance report (NCR) for evaluation by project management. Design corrections or repairs must be documented in the associated product documentation, drawings and procedures.

Non-conformance is an apparent or proven condition of any item or documentation that does not meet the specified requirements. In addition, any apparent or proven condition of any item or documentation which is considered likely to lead to incorrect operation or failure of the item or mission must be treated as nonconformance. The term **non-conformance** is also used for failures, discrepancies, defects, anomalies, malfunctions and deficiencies.

Non-conformances affecting product performance, costs, schedules or an applicable requirement are generally classified as "major."

For major NCRs leading to a permanent violation of a specified requirement, a request for waiver (RFW) must be issued and submitted for customer approval. Details must be reported to the customer together with the information that the product concerned will not fulfill the specifications in the baseline configuration of the identification documentation.

Failure management is part of the product-dedicated CM plan.

Change management controls all product changes during design development, production and qualification. Change management must be described in a company manual and in a product-specific change management plan. Each product baseline change requested or proposed by the customer, prime contractor, subcontractor or supplier has to be registered and checked formally for:

- Need/reason for change
- Description of the change
- Name of the initiator, organization, date
- Change effect (for configuration items, documents, products)
- Urgency/schedule impact
- Proposed class of the change.

If a change has been approved the appropriate consequential actions must be initiated and controlled by change management until final release of the modified items. Product assurance verifies compliance for final close-out.

Change management is part of the product-dedicated CM plan.

9.2.10 Requirement Verification

The verification of product requirements is a combined activity of project management and quality management. Verification is the controlled process for certifying specified requirements as specified in written product specifications. Together with the system specification, the following verification methods must be assigned, associated to the selected model philosophy:

- **Review of Design (RoD)**
- **Analysis (A)**
- **Test (T)**
- **Inspection (I)**
- **Similarity (S).**

Depending on the validity of a product requirement, one or more verification methods may apply. In certain cases two verification methods may apply: for example, requirement verification by "analysis" will also be verified by a test, or "review of design" will be approved by hardware "inspection."

Even complex requirements can never be verified by "similarity" or "inspection" alone. The two verification methods apply only to noncritical and standard requirements.

A product requirement is considered validated if all verification results have been documented and referenced to the requirement, and have been checked for intent. All integration and test activities have to be documented in test procedures, with the results recorded in test reports.

It is self-evident that only approved and certified facilities, workshops, clean rooms and tools may be used for such work. The verification check is documented and provided to the customer in the form of a **verification control document** (VCD).

The responsibilities for the verification checking process must be declared by project management, which is normally project management itself, system engineering and product assurance.

The verification process (Figure 9.2.11) must be traceable and documented in the form of the verification control document.

9.2.11 Lessons Learned

Quality policy, quality targets, data analyses and corrective and preventive actions have to be continuously observed to check the effectiveness of the organization's quality management system.

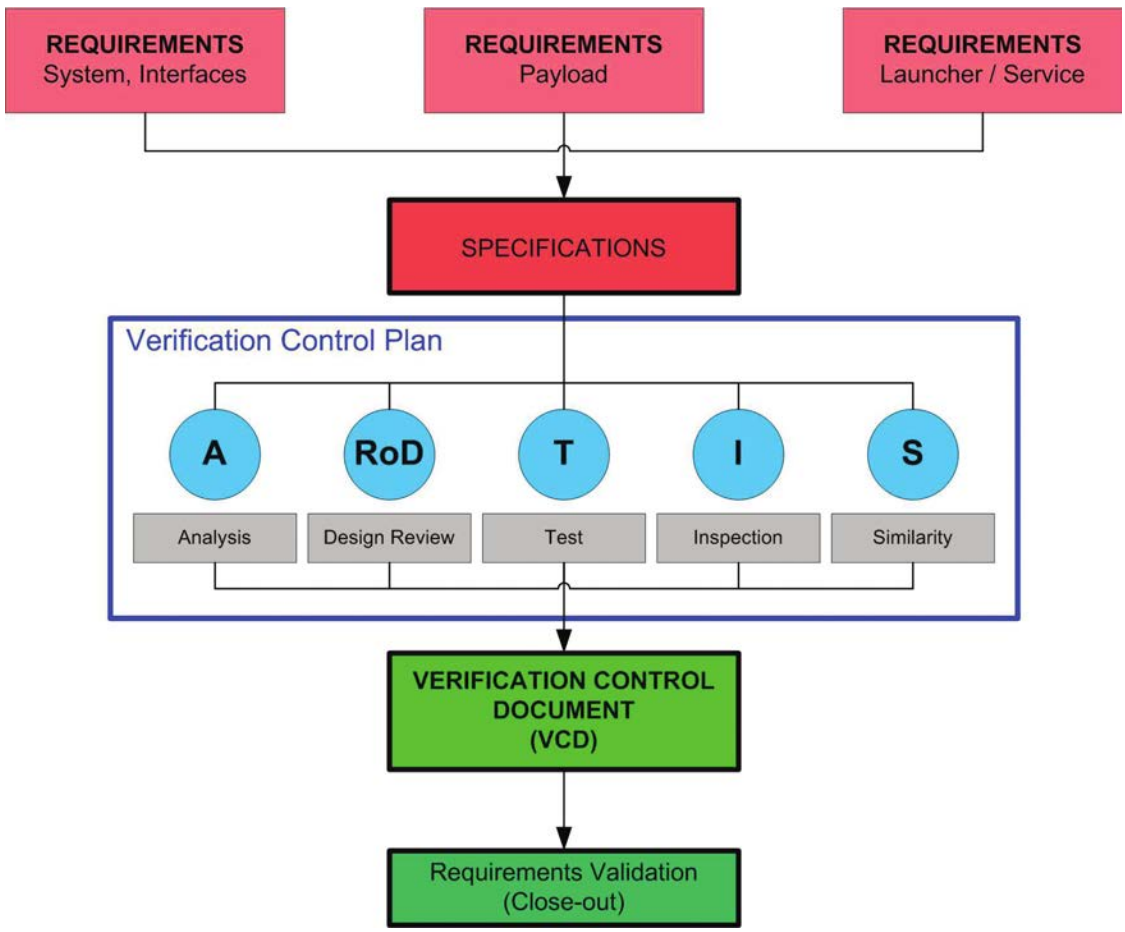


Figure 9.2.11: *Requirements verification flow.*

In order to avoid the reoccurrence of failures, product generation and project management have to be systematically analyzed. Project managers must be requested to prepare reports on special concerns and observations, accumulation of failures or management problems during the project life cycle.

Then the nature of these observations should be evaluated as to whether they are unique or systematic occurrences caused, for example, by improper tools or test equipment, or deficient instruction or education. In this case corrective actions and measures have to be initiated to avoid their recurrence in other projects. These "lessons learned" should be available to all employees involved in product generation.

9.2.12 Summary

A **qualified and certified quality management system** is the basis for quality assurance and space-related product assurance within an organization.

Universal quality management processes and procedures are described in company manuals and are transferred into product-specific PA plans.

All quality requirements as well as all product requirements have to be verified and documented to the satisfaction of the organization and its customers.

Bibliography

[9.2.1] DIN EN ISO 9001. *Qualitätsmanagementsysteme, Forderungen*. Berlin: Beuth Verlag, 2000.

[9.2.2] EN ISO 9100. *Luft- und Raumfahrt Qualitätsmanagementsysteme*. Berlin: Beuth Verlag, 2003.

[9.2.3] AQAP 2110. *NATO Quality Assurance for Design, Development and Production*. Bundesamt für Wehrtechnik und Beschaffung, Koblenz, 2003.

[9.2.4] *Konzepte und Methoden für die Gestaltung des Produktplanungsprozesses*, FQS-DGQ-Band 96-07, Band II. Frankfurt a.M.: FQS, 1997.

[9.2.5] Defense Supply Center Columbus. Military Standards. www.dscc.dla.mil/Programs/MilSpec/.

[9.2.6] AECMA, European Association of Aerospace Industries. www.aecma.org.

[9.2.7] ECSS, European Coordination for Space Standardization. http://www.ecss.nl.

[9.2.8] ECSS-M-00. *Management for Space Products*. http://www.ecss.nl.

[9.2.9] ECSS-E-00. *Engineering for Space Products*. http://www.ecss.nl.

[9.2.10] ECSS-Q-00. *Product Assurance for Space Products*. http://www.ecss.nl.

[9.2.11] ECSS-Q-80B. *Software Product Assurance*. http://www.ecss.nl.

[9.2.12] ISO 14644. *Cleanrooms & Associated Controlled Environment*. http://store.mil-standards.com.

[9.2.13] *NASA Standards*. http://pspr-pub.jsc.nasa.gov, June 2007.

[9.2.14] *NSTS, National Space Transportation System (NASA)*. http://pspr-pub.jsc.nasa.gov, June 2007.

[9.2.15] *NHB NASA Handbook (Quality and Safety)*. http://pspr-pub.jsc.nasa.gov, June 2007.

[9.2.16] *SSP Space Station Programme (Quality and Safety)*. http://pspr-pub.jsc.nasa.gov, June 2007.

[9.2.17] DLR-RF-PS-001. *PA & Safety Requirements for DLR-Space Projects*. http://dlr.de, June 2007.

[9.2.18] DLR-RF-PS-002. *PA & Safety Requirements for DLR-Space Projects*. http://dlr.de, June 2007.

[9.2.19] Managementprozesse. *HB-001-OHB OHB-Technology*. OHB-System, Bremen, 2008.

[9.2.20] Hauptprozesse Raumfahrt. *HB-001-SYS OHB-System*. OHB-System, Bremen, 2006.

9.3 Cost Management

Torsten Bieler and Sven Abitzsch

9.3.1 Introduction

Cost management (CM), also defined as **strategic cost calculation**, comprises all measures which serve to influence cost structure, cost behavior and cost level, such as the planning, estimating, budgeting and control of costs.

Early knowledge of the required resources and the available budget is an important condition for effectively influencing the product and its cost. It is therefore important to assess from an early stage the payment plans, actual expenditures, risks and opportunities. Well-implemented cost management additionally facilitates the prediction of variations as well as the definition and implementation of corrective measures to avoid cost increases. Through the forecast of payment flows, cost management smoothes the progress of budget planning and focuses the view on payment commitments.

The method of using **cost as an independent design variable** (CAIV) has a major impact on product

performance and cost efficiency. Total costs, cost trends, as well as potential risks and opportunities play an important role in the decision-making process on tactical and strategic levels.

9.3.2 Objective

The general objective of cost management is to ensure an effective and proactive project control process. Cost management establishes the overall planning baseline for work and costs over the course of a project and allocates them to the different stakeholders. This creates a common understanding among all the project participants about the costs and work effort involved.

Because of a strong interrelationship between schedule and cost management, both tasks are integrated most of the time in project control. Given this situation, the main objectives of cost management are defined as follows:

- Accurate planning and coordination of procurement activities, expenses and resources in a project.
- Detecting cost deviations during project execution and making timely recommendations for corrective actions with the aim of completing the project within the schedule and budget.

Furthermore, these objectives are to be achieved in a proactive manner, thus neutralizing negative cost trends as soon as possible. The importance of these objectives is evident when considering the typical **life cycle cost** trend of a space project, as shown in Figure 9.3.1. This figure shows that total project cost is to a large extent already determined in the early project phases. The potential to influence project cost decreases rapidly with the progress of a project, leaving less than 20% of the potential to influence when the detailed design starts. Any greater modification introduced at that point – even if intended to save cost – would result in total project cost increases and most probably would delay the project.

Therefore, cost management is most important at the beginning of a project in the mission concept phase when the greatest potential exists to effectively influence project cost. This leads to the insight that:

Cost management is more a design function than an administrative function.

Consequently, cost management is considered to be an engineering function represented by the discipline of **cost engineering**.

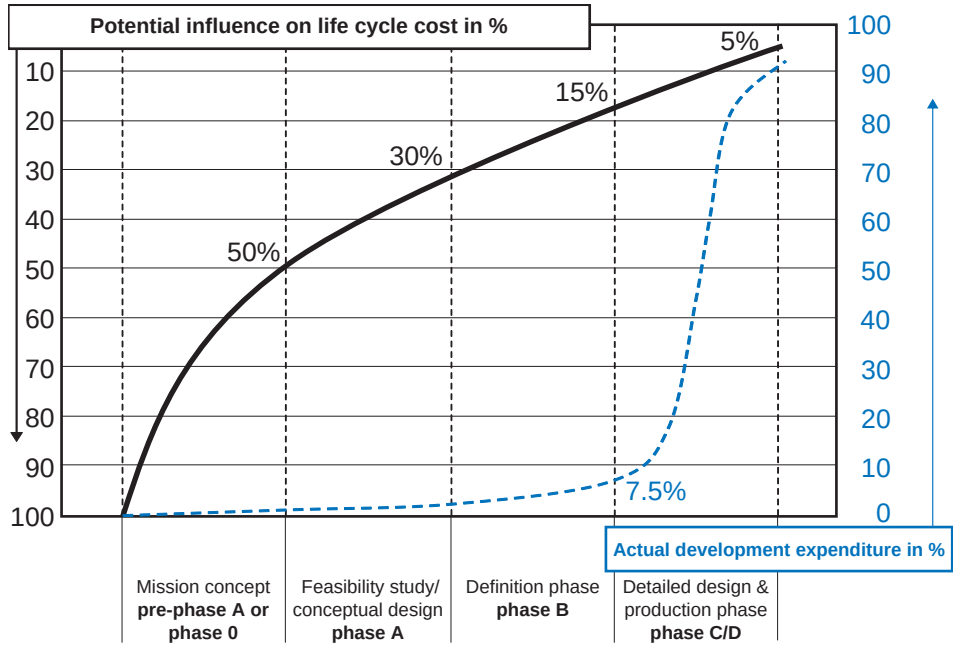


Figure 9.3.1: *Life cycle costs (commitment and cumulative) during the project phases.*



Figure 9.3.2: *Overview of the cost management function.*

9.3.3 Cost Management Process

As per ECSS-M-60B [9.3.1] the cost management task comprises the following three main activities:

- Cost estimating and planning
- Cost control
- Cost reporting.

To perform these tasks the following structures are required

- Work breakdown structure (WBS)
- Cost breakdown structure (CBS)
- Business agreement structure (BAS)
- Country/company structure (CCS).[1]

Figures 9.3.2 and 9.3.3 show the implementation of cost management as a basic function in the course of the project.

[1] This is particularly important in the ESA context since the guidelines for the geographical distribution of costs have to be respected. For non-ESA projects this is possibly unnecessary.

9.3.3.1 Process Overview

Cost management, like any other management task, applies a cyclic approach. Initial analyses/estimates of potential project costs, which are derived from parametric system cost models, are required for mission analysis and feasibility studies (Phases 0 and A). During the continuous course of the project more precise cost estimates, based on the subsystem or equipment level, are demanded. At this point more accurate cost estimation methods can be applied since more detailed project information becomes known and available. Typically, more analytical methods are considered, while the number of costing iterations depends mainly on the rate of change of the product definition. Normally a cost estimate update is required for every major project review.

9.3.3.2 Embedding into the Project Management Cycle

Part of the required cost management input is derived from the schedule management, as described in Section 9.1. These are in particular:

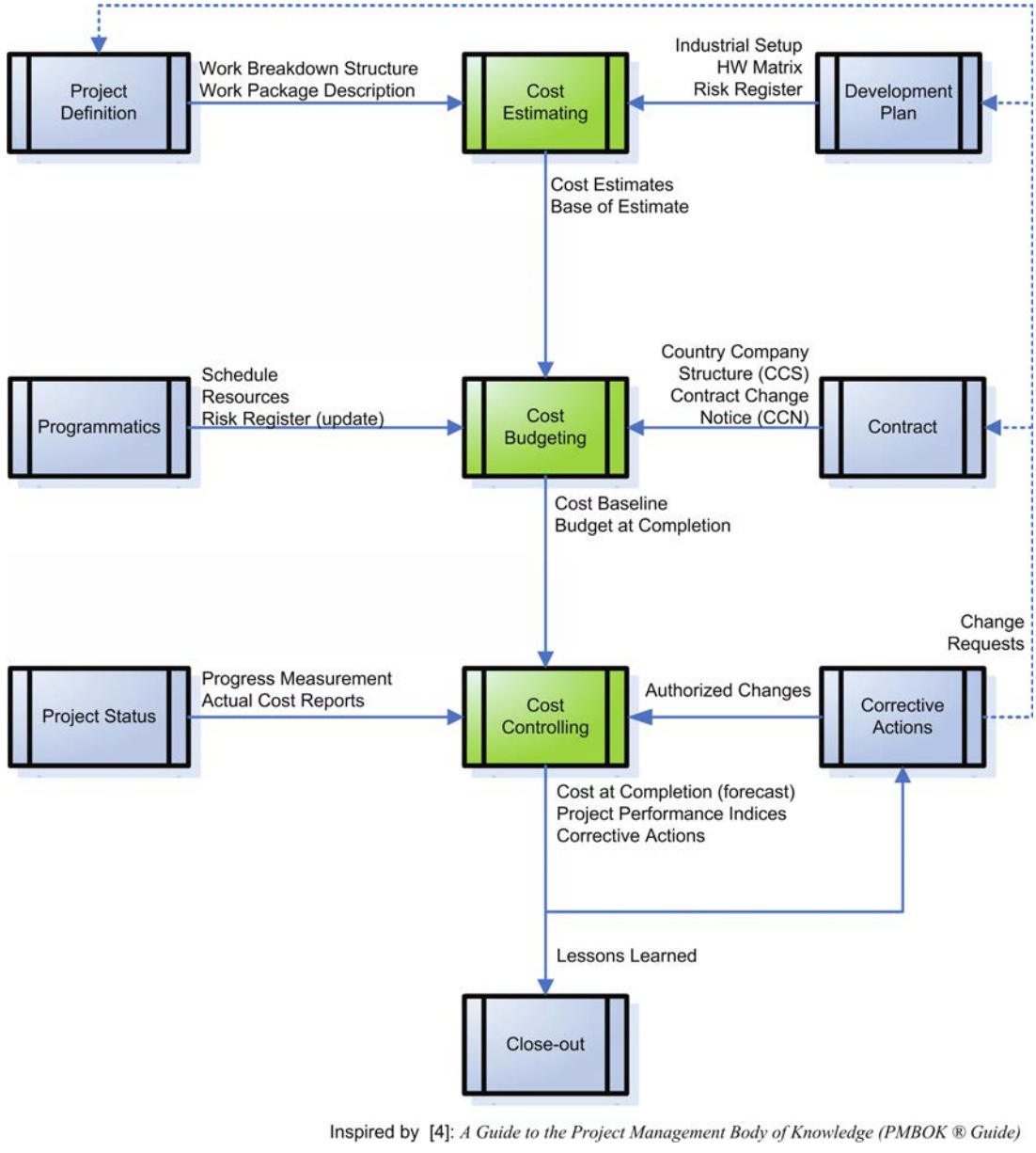


Figure 9.3.3: *Overview of the* ***cost management process.***

- Task definition
- Task duration
- Network of tasks
- Project schedule (baseline and current status)
- Resource planning.

Together with financial data from industry, **project control** provides important inputs for the cost management process. This includes contract change notices (CCNs) along with their additional cost, and project progress reports together with the running costs.

After having made **cost estimates** for all CBS elements (e.g., activities, equipment) they are arranged on the project timeline according to the schedule. This results in the project cost baseline or performance measurement baseline (BCWS) and the total project cost (cost at completion). If allocated to the country/company structure (CCS) the geographical return can be planned.

Project cost baseline and geographical return are considered the main inputs for the subsequent **cost control** process since they constitute the planned cost flows within a project. These are compared to the actual costs in the course of the project, which makes it possible to identify performance deviations and evaluate project efficiency. Performance indices are calculated and then used to estimate cost at completion (**EAC**, Estimate At Completion) and cost to complete the project (**ETC**, Estimate To Complete). Both EAC and ETC are deemed very important indicators for project management in preparing its decisions.

All these indicators, indices and reports are required to ensure systematic and robust project cost management. Figure 9.3.3 depicts how this information is embedded in the cost management process.

9.3.4 Tasks of Cost Management

9.3.4.1 Cost estimating

The responsibility for managing complex, high-tech programs which include a large proportion of new developments highlights the importance of cost and financial aspects, and in particular the difficulties associated with estimating, assessing and controlling project costs. Costs and their generation are an important factor during the entire project life cycle and play a major role when assessing different options, cost/technical trade-offs, budgets, price proposals, preparations for contract negotiations and the impact of design changes.

Considering the progress of concentration in the space industry, **independent cost estimates** (ICEs) have become increasingly important. The lack of competition results more and more often in a single-offer situation. This makes it very difficult for the customer to evaluate whether a price is justified. Obtaining an independent cost estimate can be a good basis for price comparisons (**benchmarking**) by introducing price targets (target pricing). This method provides a "second" opinion offering a more solid starting position in negotiations. Consequently, independent cost estimates are frequently used to create a situation of **virtual competition**.

All **cost elements** with their associated influencing parameters have to be identified. The required cost elements vary with the industrial branch and the item or system to be estimated. Elements typical of the space industry and their key factors or influencing parameters are listed in Table 9.3.1.

Input

Before starting the selection of the most appropriate cost estimate it has to be ensured that for each element to be estimated information is available that is as detailed as possible.

A well-populated **database** is the main basis for most cost estimating and cost engineering activities. Such a database includes cost data (e.g., cost at completion or at the time of proposal) together with the respective technical (e.g., mass, power, performance,

Table 9.3.1: *Typical cost elements for space systems and their effects.*

Cost element	Key factor/influencing parameters
Equipment	Technology, representative parameters, model philosophy
Project office	Design effort, management, quality and product assurance
Verification, integration, test	Effort for assembly, integration, verification and qualification
GSE	Machines, tools and jigs, etc.
Specific activities	Specific tests (cryosystems), planetary protection
Launch cost	Launch preparation and integration, launcher system
Operations	Operations concept, ground segment
Cost risk	Technical, programmatic, political project risks
Procurement	Georeturn, procurement policy, market situation

material, etc.) and programmatic (e.g., hardware matrix, verification and validation plan, etc.) information. Generally, proposal data as well as market studies serve as information sources. Furthermore, historic project data is taken into account if a clear record is available listing all price changes together with the associated explanation for each WBS element.

Additional information can be stored to ensure that the reference project can be understood and that important parameters influencing the estimation process, such as the industrial policy background at the time of the project, can be considered.

Other preparatory activities include a WBS analysis with respect to the project work that will have to be performed. Also, it must be checked whether any activities have been overlooked or listed more than once.

The identification of **cost drivers** is particularly important since they play a significant role in the choice of the appropriate cost estimation method and cost model to be applied.

A few important ECSS-based reference documents which should be consulted for purposes of cost control during the cost estimation process and in the course of a project are explained below.

Starting from the functional requirements in the form of a **function tree** (FT), the **project breakdown structure** (PBS)**,** derived directly from customer requirements, follows. Such a PBS is the basis for a common understanding between the different groups involved in the project as to the definition of elements as well as the related tasks and resources. Furthermore, all responsibilities within the organizations as well as coordination and optimization of the required resources and operations are described.

The **product tree** (PT), which is developed as a next step, contains the detailed definition of each system element.

As indicated in Figure 9.3.4, a description of the necessary development and production processes produces the WBS.

Work Breakdown Structure

The WBS is an effective management tool supporting both the customer and the supplier in fulfilling their business obligations, while setting the framework for regulating, planning and controlling costs, schedules and technical content.

The project is subdivided into manageable **work packages** (WPs) organized by the type of work to be performed. All activities are identified in detail. The WBS is derived from the product tree by expanding its components with **support functions** (see Figure 9.3.4) provided by the contractor. Support functions are all those activities necessary to produce the end product (e.g., project management, engineering, product assurance support). It should be mentioned that an end product can be a development, hardware (e.g., a flight model) or a service (such as mission operations).

Work Package Description

The WPD is closely connected to the WBS. This directory contains specifications for each work package, its title and unique code, as well as precisely defined tasks, deliveries and expected results. According to ECSS-M-10, the following information should be addressed:

- Project name and phase
- WP title
- Unique identification of each WP (for each project phase) in line with the identification rules
- Supplier or entity in charge of the WP
- WP manager's name and organization
- Supplier's country (in case political or economic constraints exist)
- Product to which the tasks of the WP are allocated
- General description of the objective of the WP
- Detailed description of specific tasks, including cost categories (see also ECSS-M-60) and type (product related, expense related, supply related)
- List of the guidelines to be followed in performing the task
- Interface links with other tasks or WP required to set up a planning network as per ECSS-M-60
- List of constraints, requirements, standards, and regulations
- List of the expected outputs
- List of deliverables
- Location of delivery

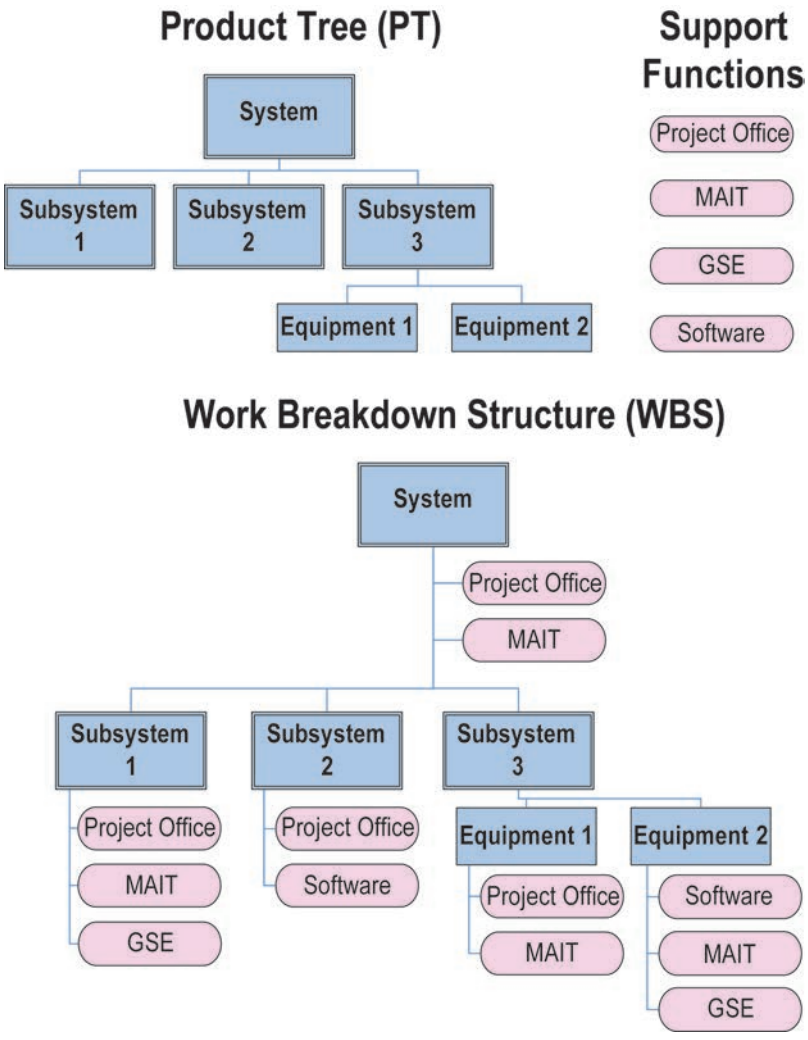


Figure *9.3.4: Example for developing a work breakdown structure (WBS).*

- Start event identification including date
- End event identification including date
- Excluded tasks.

Furthermore, those tasks that are explicitly excluded should be mentioned.

Cost Breakdown Structure

The CBS defines a set of cost categories used to allocate all the costs of the project. It also provides a common frame for all cost management activities of all parties involved in the project. The total cost planned for each WP is divided by cost category, such as labor, material or subcontractors. Each supplier distinguishes furthermore between direct (labor, facilities, etc.) and indirect (common) costs for each cost category.

Business Agreement Structure

The BAS identifies the project responsibilities and relationships between the customers and suppliers by relating WPs within the WBS. Furthermore the BAS supports the cost management process by providing

the means to properly assess the impact of potential modifications and changes (e.g., in liability, financing, technology or schedule) at each level of the contractual hierarchy. It is applied to trace contractual responsibilities and interface definitions, and if required can also be used to verify and monitor the **geographical distribution** specified in the country/company structure (CCS) in the case of international projects.

Risk Register

As described in Section 9.1.4.3, potential risks to the project are listed in the **risk register** and ranked by their impact, which is the product of the probability of occurrence and potential consequences. This information is very important for cost management since it is the basis for specifying the project cost risk and deriving a safety margin for the project budget.

Risk management, which is closely connected to project and cost management, has developed during the last few years into a specific function for systematically identifying and analyzing project risks.

Risk factors of particular interest for cost management are for example programmatic parameters, implementation of new technologies, specific cost drivers, financial (and eventually georeturn) conditions.

Project Schedule

Cost and schedule management are closely linked in the task of deriving a network plan from the WBS. The network plan connects all activities with their logical relations in a time sequence. Milestones, defining start and end dates and/or required design reviews, are introduced into the timeline to provide a better overview. Identifying the critical path helps to anticipate developments which may require appropriate corrective measures. Project timelines are generally displayed in a Gantt chart in which each activity is represented by a bar. The bar lengths indicates the duration and the logical relationships are symbolized by arrows.

Economic Conditions

The economic conditions can be understood as relating to a point in time to which all financial specification and information are compared. It is, so to speak, a norm and is important for calculating inflation costs at a later date. The information is typically put in the following form: e.c. month year (e.c. Jan 2009).

Technology Readiness Level

The **technology readiness level** (TRL) serves as an indicator of technology maturity. It can be used to determine the following parameters, among others:

- To what extent are technologies, materials, tools, equipment, etc., already developed and available?
- Which qualification tests have been already performed and which are still outstanding?
- What is the expected timeline?
- To what order of magnitude do modifications have to be expected?

TRLs are taken into account when assessing potential risks concerning the costs and schedule. The generally applied TRL definitions are based on a NASA White Paper from 1995 [9.3.5].

Cost Estimation Requirements

All of the above-mentioned information necessary for the cost estimation process is specified in the invitation to tender and included in the list of **deliverables**.

All the data and information mentioned in the sections above form the basis for selecting the most appropriate cost estimate methodology. The choice of cost models and how they are constructed depends very much on the basic assumptions, the procurement of adequate input data, the current and predicted project phase(s), as well as suitable references. References are stored in specific databases and include relevant technical, programmatic as well as cost allocation information.

Different **cost estimation methods and models** are described below.

Methods/Models

There are different approaches and methods for estimating and assessing costs, all of which have advantages and disadvantages under particular circumstances. The most suitable methods are selected by considering the following:

- Character of the activity to be costed or assessed.

- The organization's experience with the system or activity to be costed
- The extent to which reference can be made to previous exercises
- The availability of reliable design information
- The time available to prepare the estimate.

Other selection aspects are the project phase in which the estimate is to be performed and specific requirements concerning the cost details as well as the degree of accuracy required. Usually several methods are applied as a "plausibility check" to verify that the results are valid. The most common approaches for estimating and assessing costs are presented next.

Expert's Judgment or Wideband Delphi Method
This approach is used to derive a rough and rapid estimate of the order of magnitude of a cost. Its relevance is normally limited to specific areas and implies expert judgment as well as close familiarity with the field of activity. Such a "rule of thumb" approximation is not very sophisticated, but it may be sufficiently accurate in certain circumstances when an estimate is required quickly.

Bottom-up Cost Estimate
This method derives detailed cost estimates at a relative low work package level of the WBS and is closely connected with schedule planning and resource allocation. Due to the amount of detail, a bottom-up estimate is both time consuming and expensive. A precondition for this method is good knowledge and an appropriate level of definition of the activities that have to be estimated.

Analogy
An important condition for this commonly applied method is being able to ascertain the cost of previous activities, elements or systems. Relevant historical data is used as a reference for predicting the cost of current items. The method depends very much on the completeness and accuracy of the available information and on the extent to which technical, programmatic or economic differences between the reference and the element to be estimated can be identified and utilized to normalize the database. Also, cost trends and any changes in circumstances that might influence the costs must be taken into consideration.

Competitive Supplier Proposals
Wherever it is intended to subcontract an activity, binding subcontractor proposals probably provide the most reliable estimate. However, each customer should preserve the capability to estimate and assess the cost of activities to be subcontracted. This is especially true whenever there is only little or no real competition, or when subsequent customer-generated modifications are likely.

Parametric Analysis
Parametric analysis requires careful consideration of costs as well as technical and programmatic information. This makes it possible to identify cost drivers and to develop cost models. This approach effectively correlates labor costs and personnel with parameters describing the product to be costed. This correlation leads to regression formulas known as **cost estimation relationships** (CERs). CER can range from simple mathematical relationships to highly complex equations and can be drawn up wherever there is an adequate amount of data available. CERs are typically developed for a particular technology taking into account associated representative parameters (such as mass, power or surface) and the model philosophy. The following equation is an example of a simple CER:

$$\text{Costs} = a \cdot P^b \cdot n$$

where parameters a and b are determined by statistical analyses of cost reference data. In this case the technical parameter is P, for example the unit mass, and n could be the number of models to be developed.

Parametric cost models can be either developed within an organization to analyze its own data or purchased on the commercial market. In the latter case it is essential to calibrate the models with specific reference data from the user's organization. Even though there may be significant costs associated with developing such models or purchasing licenses for commercial versions, they have a number of advantages. They allow estimates or assessments to be made fairly rapidly and at relatively low cost, the evolution of the processes involved can be traced, and they can be rerun at a later point in time.

It is possible to bring together different methods for different cost elements and to merge and develop them further into a specific cost model, for example for optical instruments. Such models might then combine general design parameters such as aperture sizes, wavelengths or the type of telescope with more specific parameters such as mass or development effort.

For every method and model the following should be considered:

- Application limitations of the models
- Sensitivity of certain parameters
- Contract and economic conditions
- Impact of the industrial consortium
- Need for a plausibility check of the results
- Best opportunity for applying the model results.

Output

Last but not least, cost estimates serve the customer as well as the client as a common basis for understanding. A cost estimate as a quantitative evaluation of the most likely project cost can be presented either in summarized or detailed form. As described in Table 9.3.1, a cost estimate contains a variety of cost categories such as labor, material, subsystems, services, facilities and information technology, as well as specific categories such as allocations for inflation or reserves. The extent and type of additional, supporting information depend on requirements and the field of application. Independent of the level of detail, the supporting documentation should provide a clear, professional and complete picture of how the cost estimate was derived. The supporting documentation should include the following:

- Project content description
- Description of the assumptions
- Description of the estimation
- A cost risk/cost opportunity analysis
- Trade-off analyses
- Description of constraints and boundary conditions
- Definition of the cost estimation process; that is, the method used to implement the cost estimation process in the organization vis-à-vis the subcontractors.

9.3.4.2 Cost Budgeting

Cost budgeting is the process of planning cost for a project by aggregating cost estimates for single work packages (WPs) into a binding **cost baseline**, thus defining the cost flows planned for the duration of the project.

Input

The following information and input are required for the cost budgeting process:

- Work breakdown structure (WBS)
- Work package description (WPD)
- Cost estimates (e.g., for WPs, subsystems, tasks, assemblies, material, etc.)
- Basis of estimate (BoE) (i.e., documentation of the assumptions behind the cost estimates)
- Project schedule
- Availability of resources (e.g., resource calendar, cost limits, financial planning)
- Contract information (e.g., prices for external products and services).

Methods

Cost budgeting can be broken down into three major steps, described as follows:

1. **Aggregation of the Cost Estimates:** Allocation of all estimated cost items to WPs according to the project's WBS. Aggregation of these costs to higher nodes of the WBS, resulting in the total project cost at the top WBS node. In addition, by taking into consideration the schedule information for all WPs, the project's **cost profile** (planned spending over time) can be derived.
2. **Establishing a Management Reserve:** The inherent cost risk of a project – analyzed as part of the cost estimation process – is reflected in a management reserve incorporated in the project budget. This reserve can be tapped to fund potential mitigation actions which may become necessary in the course of a project to avoid risks or to reduce the impact of emerging risks. It should be noted that the reserve is not part of the cost baseline, meaning that it is not allocated to specific WPs, and that it can only be used with the authorization of the project manager or customer. In recent cost risk analyses the overall cost risk of a project is broken down into the hierarchy levels in a project organization which represent the risk source. In this case the project's management reserve covers only the cost risk contributions directly associated with carrying out

a project (e.g., technical development, planning, scheduling). Other cost risk contributions, such as more indirect events stemming from the project environment (e.g., political and market risk) and *force majeure*, are typically borne at the corporate level (i.e., by the company or space agency) and are thus not budgeted to the project.

3. **Smoothing the Cost Profile:** The original cost profile derived from steps 1 and 2 needs to be adjusted to when financial resources become available. The objective is to facilitate funding by achieving a smooth cost profile which avoids extreme fluctuations and remains within the customer's affordability limits. In the case of development projects, the best practice regards an S-shaped cumulative cost curve as being the most favorable profile for total life cost, balancing the project's duration and cost. Besides utilizing financing instruments to decouple spending from cost flows, a shifting of WP tasks or schedules might also be required to smooth the cost profile. It should be noted that such activities generally have a detrimental impact on project costing and scheduling. Moreover, the practice could even introduce new conflicts with regard to resource availability and milestones. As such, cost budgeting for a project is an iterative process eventually leading to the project's cost baseline.

Output

The main outputs of cost budgeting are:

- **Cost Baseline:** Planned cost for scheduled work over the project's timeline. It is used as a reference for cost and schedule performance measurement.
- **Funding Plan:** Date and amount of funding steps to finance the project costs, aligned with the cost baseline.
- **Change Requests (CRs):** The cost budgeting process could result in change requests compared to the project schedule. After being assessed and accepted, such changes have to be documented in adjusted management plans (this is the task of change management).

9.3.4.3 Cost Control

Cost control in general can be characterized as the process of influencing cost driving factors in such a way that cost increases are contained and opportunities for savings are exploited. In addition, cost control comprises the task of tracking and documenting all changes in the cost baseline.

The nature of the cost control process also varies with the contract price type, with the distinction usually being between a fixed price contract and a cost reimbursement contract. Their main differences can be described as follows.

In the case of a **fixed price contract** – which is the standard contract type for space projects in Europe – the product is exactly defined in scope, deliverables and WP effort. This allows a price to be fixed for the end product which includes all margins for profit, escalation/inflation and project risks. If a fixed price needs to be adjusted in the course of the project, the contract has to be changed. This can be requested by either party through a change request (CR), which must be negotiated and formally agreed by all parties. To become effective the changes have to be documented in a so-called contract change notice (CCN) which amends the original contract.

In a **cost reimbursement contract** the contractor is compensated for the expenses incurred while completing agreed-upon work. The contractor also receives a negotiated profit according to the contract specifics (fixed profit fee or profit percentage). Project risks and inflation are typically borne and paid by the customer. Consequently, for this contract type the cost per product can grow considerably since the effort to achieve a result is reimbursed, in contrast to payment for a completed product in fixed price contracts. On the other hand, this contract type allows adjustments in the project plan to be easily handled by redirecting effort among WPs. If required, the customer is able to exercise direct control over the work program through this mechanism. This flexibility in project execution renders cost reimbursement contracts very suitable for major development or state-of-the-art projects, where typically neither the product nor the work plan can be defined up-front in detail and changes are likely.

In view of cost control, cost reimbursement contracts may seem to be disadvantageous, as they bear a higher potential of unwarranted payments which may lead to cost overruns. Additionally, all adjustments to the original baseline have to be assessed with respect to their impact on schedules and cost at completion so that budgets are not exceeded or delays introduced.

Accordingly, several precautions are necessary on the customers' side governing the cost control process in cost reimbursement contracts. First, cost targets per WP are set and scheduled on the project timeline. Even though these targets are not contractually binding, they are utilized as a cost and performance measurement baseline. Second, the customer needs to set up an appropriate progress tracking system to trace accomplished work. In combination with the performance measurement baseline this allows assessment of the justification for costs claimed by contractors. In particular, such a system lays the foundation for the **earned value management** to derive cost and schedule performance indices and to facilitate trend analysis regarding cost at completion, further explained under "Methods" in the section below. Moreover, cost discipline should be enforced by bonus agreements for exceeding project performance (e.g., early delivery) and by penalties for underperformance.

Cost reimbursement contracts are the standard contract type in the USA for aerospace and government projects.

Input

For both contract types the following input is required:

- Cost baseline
- Funding plan
- Project progress reports, documenting work progress and deviations with respect to the project baseline (cost and schedule)
- Project management plan (PMP), which in this case specifies the scope and responsibilities associated with the project control process.

Additional input for projects on fixed price contracts:

- Payment plan (payment milestones)
- Contract change notices (CCNs).

Additional input for projects on cost reimbursement contracts:

- Authorized changes to the cost baseline, contract, scope of work, project schedule.
- Performance reports at the cut-off date of a specified reporting period. In particular the following reports are required:
  - Percentage of work complete (physical progress) for ongoing WPs and activities
  - Actual cost (authorized cost incurred) at cut-off date.

Methods

The key to cost control in all contract types is the definition of the **percentage of work complete**. It determines the actual work finished in a WP versus the overall planned work for that WP. Even though this definition is sound and clear there are some practical problems involved in measuring appropriately the finished work portion. For example, many projects determine the percentage of work completed via the ratio of work hours spent over hours scheduled. This appears reliable at first sight but, as is commonly known, effort is not strictly correlated with progress, which may lead to skewed reporting. Even worse is the common practice of taking the task manager's subjective judgment on the achievement as input for work complete. This bears a high risk of biasing the report, since even the most experienced manager is optimistic about assessing the work status. Moreover, the perspectives on achievement may diverge dramatically between customer and contractor. Lessons learned confirm that such methods do not qualify for rigorous cost control as they are systematically inaccurate because of deviations between status report and reality.

To overcome this problem, a reliable and objective method for determining work progress must be found. The solution is to introduce **physical progress measurement** for each WP that is to be monitored for cost control purposes. With physical progress measurement the percentage of work complete is predefined and linked to physically measurable achievements (i.e., contractually agreed deliverables or events). With this method a rigorous progress tracking system can be established throughout the project. For space projects the delivery of development models turned out to establish a good physical measurement leading to a typical scale of completion percentages as follows: 10% for structural model delivery, 85% for prototype delivery and 100% at acceptance review sign-off.

Cost Control in Fixed Price Contracts

In contrast to popular opinion, a fixed price contract does not protect from price increases. Therefore,

rigorous cost control based on progress measurement is also required for fixed price contracts. When scrutinized, the term **fixed price** actually means a certain price for a fully specified product (work program and deliverables). In other words, changing the product specification – even if required for a good reason – formally leads to renegotiation and potentially to price increases.

Considering space projects, in mostly new developments or state-of-the-art projects trying to advance technological limits, the specific definition of a product has to be vague by nature. This prepares the ground for repeated product adjustments, change requests and CCNs with their cost increases. It is obvious that such an environment requires good and systematic cost control.

Generally in fixed price contracts, payments are linked to project milestones, forming thereby so-called payment milestones. Typically, project reviews (SRR, PDR, CDR, etc.) or handovers of deliverables are best suited to serve as payment milestones because they represent the finalization of a dedicated project phase. The project costs are determined for these project phases using the cost baseline, and the respective amount of the scheduled cost is allocated to the payment milestone representing that phase. This amount may need to be supplemented with procurement costs for long-lead items requiring advance purchase.

In this setup the main task of cost control is to substantiate that physical progress at the payment milestone is in line with the schedule, thus justifying the payment of the predefined amount according to the cost baseline. If any deviations occur, the cost controller has the task of identifying corrective actions and reporting them to project management for further implementation. Other tasks include the assessment of contract change requests (CCRs), preparing the respective negotiations, documenting cost changes due to CCNs, and continuously updating the estimated cost at project completion, keeping the project manager informed of the overall project cost whenever a change is made.

Cost Control in Cost Reimbursement Contracts

In this contract type the contractor is entitled to timely compensation for costs incurred. In cases of low work efficiency this holds the danger of an imbalance between work performance and incurred cost. In the long run this leads to overpayments, which add up to higher cost at completion. Because the cost at completion reacts quite sensitively to fluctuations in contractor efficiency, the most important task for the cost controller is to detect as early as possible any deviations from the cost baseline, then to identify the cause, quantify possible effects – especially on cost at completion – and assess whether corrective action needs to be taken.

In **earned value analysis** (EVA) a well-proven performance measurement method exists to assist the cost control process in its task of identifying deviations. The purpose of EVA is to determine at a predefined cut-off date (e.g., the end of the month) the value of work performed. This makes it possible to compare in an integrated way scheduled work, accomplished work and actual effort. In order to carry out such an integrated assessment all these parameters need common metrics. They are therefore expressed in cost units.

The EVA method of performance measurement can be scaled to any project level, namely single activities, WPs or the entire project. In large projects it is best practice to limit the effort for the controlling process by selecting a typical subset of work packages on which EVA is carried out. These are designated as **control accounts** (CAs). They are also used to delegate responsibility in the controlling process to different contractor levels by designating control account managers (CAMs). The three control values required for EVA are determined for every CA at every cut-off date. They are described as follows:

- **Budgeted Cost of Work Scheduled (BCWS):** Planned cost for work (activity, WP, project) at cut-off date according to the cost baseline.
- **Budgeted Cost of Work Performed (BCWP)** or earned value: Budgeted cost for the work (activity, WP, project) actually accomplished at the cut-off date. This value generally is calculated as the percentage of work complete multiplied by the budgeted total cost for a work unit:

BCWP = % work complete × budgeted total cost

- **Actual Cost of Work Performed (ACWP):** Actual cost incurred for the work (activity, WP, project) accomplished at cut-off date.

With these values, project efficiency, schedule performance and cost performance are easily determined by calculating the following indices:

- Schedule performance index (SPI) = BCWP/BCWS > 1 is good (ahead of schedule)
- Cost performance index (CPI) = BCWP/ACWP = 1 is good (cost right on plan)

The performance indices influence the originally planned project cost, denominated budget at completion (BAC). An estimate of the total project cost – the estimate at completion (EAC) – related to SPI and CPI is given by the following equation:

Estimate at completion (EAC) = ACWP + (BAC – BCWP)/(CPI × SPI)

The variance at completion (VAC), which defines the delta cost at completion of the project, is determined by the difference between the original budget and the EAC:

Variance at completion (VAC) = BAC – EAC
(a positive value is favorable)

These forecasts, performed at every cut-off date (typically on a monthly basis), provide a clear picture of the performance trend in a project. A comparison of BAC and EAC provides transparency about the most likely cost overrun if no corrective actions are taken.

Another very important parameter for project management is the cost required to complete the project or the **estimate to complete** (ETC). This information can be computed via the following relationship:

- Estimate to complete (ETC) = EAC – ACWP

If the project manager aims at completing the project within the original cost budget BAC, then the efficiency required to do so is calculated as the ratio of *remaining work* divided by *remaining budget*:

- To-complete performance index (TCPI) = (BAC – BCWP)/(BAC – ACWP)

For example, if 25% of the work has been performed with 50% of the budget (CPI = 0.5), then the subsequent TCPI has to increase to 1.5.

The difference between the TCPI and the demonstrated CPI is a good indicator for judging whether the planning is realistic. In the example given it is very unlikely that the efficiency can be improved from 50% to 150%, rendering the planning unrealistic.

A common way to communicate the findings of an EVA is to display the cost baseline (BCWS), the actual cost (ACWP) and the earned value (BCWP) as cumulative cost curves. This way the performance history, the schedule and cost variances, BAC and EAC can all be clearly displayed (see Figure 9.3.5).

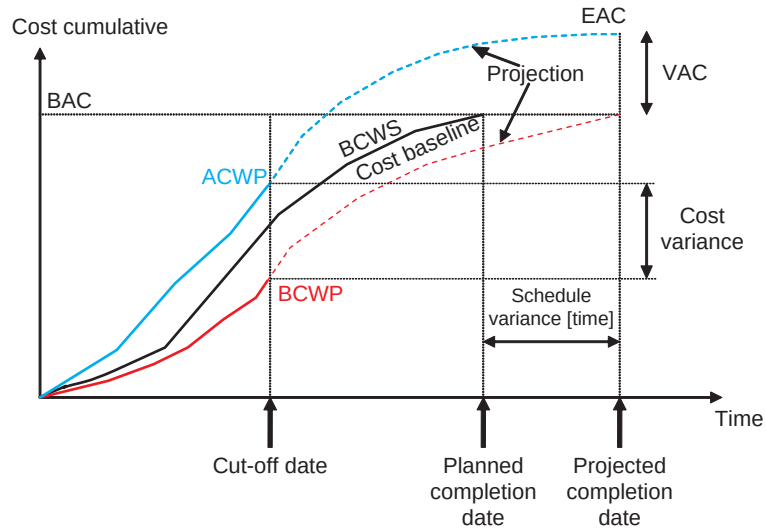


Figure 9.3.5: *Earned value analysis.*

Progress measurement and reporting are the keys to successful project cost control. They require answers to the following questions:

- What results/deliverables were planned?
- What effort was planned/budgeted for achieving the results?
- To what degree were the planned results realized?
- Which effort was required to achieve the results?
- Which effort should have been claimed for the results (→ earned value)?

Output

The following information is typically the expected output of a cost control process:

- Modification of original cost estimates and modification of the documentation about the basis of the estimate (BOE)
- Modification of the cost baseline according to authorized changes
- Performance indices for control accounts and specifically selected elements of the WBS
- Status reports on the project progress (physical progress measurement of control accounts)
- Estimate of total project cost at completion (EAC)
- Projection of the most likely completion date based on demonstrated schedule performance
- Recommendation of corrective action to improve project efficiency (e.g., budget adjustments, reallocation of resources, reorganization of project schedule/network)
- Modification of the project management plan (reorganization)
- Documentation of "lessons learned" (analysis of root causes for cost deviations and their documentation).

9.3.5 Close-out

The end of a cost management process is characterized by the so-called **project close-out**, documenting lessons learned and formulating best practices for coming projects. This last phase of the cost management process is often neglected, mostly for organizational reasons. The project team is dissolving, new tasks lie ahead and obligations to exchange information may have ceased with the end of the contract.

It should, however, be noted that a proper project close-out is of utmost importance for cost management, since all processes, methods and models need verification, calibration and feedback. Only if experiences are recorded, documented and exploited are future improvements likely.

Therefore, the task of cost management after completing a project lies in a comparison of the actual cost at completion (CAC) to former budgets and estimates, followed by a discrepancy analysis in case major differences are found. Actual cost and findings resulting from this analysis are generally required to calibrate cost estimating models in order to improve their accuracy.

In addition, the discrepancy analysis is also used to identify causes for cost increases in a systematic manner. Causes for cost increases are typically stochastic (e.g., test failure, accident, loss of contractor) or systematic (e.g., late change in requirements, adverse market effects and procurement policies, or underestimated complexities). In the latter case they should be documented as lessons learned and communicated to the project organizers so they can be avoided in the future.

9.3.6 Outlook

9.3.6.1 Virtual Competition

The general trend to mergers in the aerospace industry leads to conglomeration and less competition in this industrial sector. There are already some areas, especially in the space business, where only one supplier exists, posing a risk of monopoly and high prices detrimental to space projects.

Utilizing the cost estimating capacity of good cost management could help to overcome such detrimental monopoly situations. The idea is to confront a monopoly supplier with a robust cost estimate for its product which urges the supplier to defend its price. This is believed to instigate the effect of price competition, possibly leading to lower prices. This approach is tagged "virtual competition."

9.3.6.2 Electronic B2B

Because of the increasing relevance of electronic procurement between business partners – called

business-to-business (B2B) procurement – cost management and parametric cost estimation methods are progressively gaining in importance. The reason is the ability to respond quickly with a quotation to an announcement of opportunity/RFQ.

If parametric cost models are harmonized and audited between the customer and the supplier, they make feasible both real-time cost proposals by the suppliers and real-time proposal evaluation by the customer. In other words, the time and cost involved in presenting and evaluating a proposal could be dramatically decreased. In addition, useless discussions about different perspectives on determining work effort could be avoided because the whole proposal is based on objective, physical design parameters (mass, thrust, data rate, area, etc.).

Such a system would be an extremely valuable asset for space agencies willing to introduce electronic procurement portals, copying the trend in the automotive industry.

Bibliography

[9.3.1] ECSS-M-60B. *Space Project Management – Cost and Schedule Management*. Noordwijk: ESA Publications Division, June 2006.

[9.3.2] ECSS-M-10B. *Project Breakdown Structures*. Noordwijk: ESA Publications Division, June 2003.

[9.3.3] ECSS-Q-00A. *Space Product Assurance*. Noordwijk: ESA Publications Division, April 1996.

[9.3.4] *A Guide to the Project Management Body of Knowledge* (PMBOK® Guide), Third Edition. Denver, CO: PMI, 2004.

[9.3.5] Mankins, J.C. *Technology Readiness Levels – A White Paper*. NASA Advanced Concepts Office, Office of Space Access and Technology, April 1995.

[9.3.6] Greves, D., Joumier, H. What is Cost Engineering and Why is it Important? *ESA Bull.*, 115, August 2003.

9.4 Legal Aspects of Space Activities

Michael Gerhard

Legal aspects need to be considered for every space activity. They arise not just when there is an incident or disagreement among the parties. Legal aspects can also govern the mission concept and should therefore be considered at a very early stage of mission planning.

In contrast to many other legal fields, the source of space law is in public international law. Public international law is in particular the law governing the relations among states as well as the law governing international organizations. Consequently, public international law does not address the individual, in principle. Nevertheless, knowledge of it is indispensable for understanding the rights and obligations of the individual, for example why space activities require authorization, or what contents a contract between the parties to a space mission should have. Public international law on space activities will therefore be introduced first (Section 9.4.1).

This will be followed by an overview of the laws directly relevant for the individual (Section 9.4.2). Finally, recent developments in space politics and space policies in the European Space Agency (ESA) as well as in the European Union (EU) will be presented.

9.4.1 Basic Rules of Space Law

Space law is based on public international law. Ninety-eight states have agreed that outer space is not subject to national appropriation. It is therefore regarded as international territory, similar to the high seas and Antarctica; no state is allowed to claim sovereignty; no national legislation applies in outer space unless agreed by public international law.

A prominent role within public international law on outer space is taken by the international treaties drafted under the auspices of the **United Nations**. These treaties were then submitted to states for signing. There are also other agreements between states which apply to outer space activities.

9.4.1.1 International Outer Space Law of the United Nations

The Committee on the Peaceful Uses of Outer Space of the United Nations has drafted five treaties on outer space since 1962 [9.4.1]. They have been signed and ratified by states as follows:

- Outer Space Treaty (98 ratifications and 27 signatories)
- Rescue Agreement (88 ratifications and 25 signatories)
- Liability Convention (82 ratifications and 25 signatories)
- Registration Convention (45 ratifications and 4 signatories)
- Moon Treaty (11 ratifications and 5 signatories).

Moreover, the General Assembly of the United Nations has also adopted seven resolutions on outer space.

Freedom to Use and Explore Outer Space and its Limits

Most relevant is the treaty on the principles governing the activities of states in the exploration and use of outer space, including the Moon and other celestial bodies, as opened for signature on January 27, 1967 (**Outer Space Treaty**).

Art. I (2) of the Outer Space Treaty guarantees the freedom of exploration and use of outer space, which is to be free and without discrimination on a basis of equality and in accordance with international law. Art. II of the Outer Space Treaty defines the limitations: outer **space** is not subject to national appropriation by claim of sovereignty, by means of use or occupation, or by any other means.

Bridging this tension, the Outer Space Treaty establishes **obligations for** states undertaking activities in outer space: The exploration and use of outer space is to be undertaken:

- for the benefit of and in the interest of all countries
- in accordance with international law, including the Charter of the United Nations
- without placing in orbit any objects carrying nuclear weapons or weapons of mass destruction
- for peaceful purposes
- with due regard to the corresponding interests of other states
- while avoiding harmful contamination of outer space and adverse changes to Earth's environment
- while opening all stations, installations and equipment on the Moon and other celestial bodies to representatives of other states on a reciprocal basis.

In addition, the Outer Space Treaty establishes an obligation to assist astronauts in the event of accidents and incidents as well as an obligation to return to the state of registry any objects which return to Earth from space [9.4.2], [9.4.3].

State Responsibility for Activities in Outer Space

According to Art. VI of the Outer Space Treaty, states are **responsible** for national activities in outer space. This responsibility exists regardless of whether the activity is undertaken by a governmental agency or by a non-governmental entity. This responsibility aims at assuring that all national activities are carried out in conformity with the Outer Space Treaty, especially with the above-mentioned obligations. For governmental activities this can be assured through internal guidelines which are also applicable to activities of the respective national space agencies. Additional regulations are required for activities undertaken by non-governmental entities. Therefore, the states have the obligation, according to Art. VI (2) of the Outer Space Treaty, to authorize and continuously supervise such nongovernmental activities in outer space. Based on this, a dozen states have enacted national space legislation which specifies authorization requirements and procedures as required by Art. VI (2) of the Outer Space Treaty; see Section 9.4.2.1 below.

Registration of Space Objects

The Convention on Registration of Objects Launched into Outer Space was opened for signature on January 14, 1975 (**Registration Convention**) and establishes a two-tiered **registration** system [9.4.4], [9.4.5], [9.4.6].

Art. II of the Registration Convention obliges the state to maintain a **national registry** of objects launched into outer space. In this registry, the state must enter information about the space objects for which this state is the launching state. A **launching state** is a state which has launched or procured the launching of the space object or from whose territory or facility a space object was launched, see Art. I (a) Registration Convention. Where there are two or more launching states for any such object, they jointly determine which of them has to register the object, Art. II (2) Registration Convention. There is only one state of registry for a space object. Only this state is subject

to the rules, which will be described below (the obligation to inform the Secretary General of the United Nations, the right to exercise jurisdiction and control). In case the space object is operated by a private legal entity, a university or a research organization, the state must be informed of the forthcoming launch so that it can enter into relevant agreements with other launching states.

Art. IV of the Registration Convention obliges the state of registry to furnish to the Secretary General of the United Nations as soon as practicable the following information: name of the launching state or states, an appropriate designation for the space object or its registration number, date and territory or location of the launch, basic orbital parameters (including nodal period, inclination, apogee and perigee) and the general function of the space object [9.4.7], [9.4.8]. The Secretary General of the United Nations, assisted by the United Nations Office for Outer Space Affairs, enters this information in the **international register** on space objects [9.4.9].

While the contents of the international register are defined by Art. IV of the Registration Convention, the contents of the national registries (and the conditions under which they are maintained) may be determined by the state of registry. It is obvious that the national registry can serve as an entry point to gather the information which must be furnished to the Secretary General of the United Nations. Additional information (such as the name of the operator, mass–density ratio) may, however, be required.

Finally, Art. VIII (1) of the Outer Space Treaty allocates **jurisdiction and control** over a space object to the state on whose national registry the object is carried – even though the object is then located in the international territory of "outer space." This guarantees that national legislation remains applicable to the space object [9.4.10].

State Liability for Damage Caused by Space Objects

States bear international **liability** for damages to another state party or to its natural or juridical persons caused by space objects, Art. VII Outer Space Treaty [9.4.11], [9.4.12]. This article provides – in case of damage – at least one liable state to the injured third party [9.4.13]. State liability is not based on the causation of damage. It is rather based on the participation of a state at the moment of launch; that is, that state has contributed to placing the object into outer space (**launching state**). It is sufficient that a state has authorized (or was obliged to authorize) the activity for which the space object is used [9.4.14]. For the state this entails liability even if the damage was caused by private activity. Therefore, the state must ensure that it receives information about all activities carried out from its territory or by its natural or juridical persons. At best the state will scrutinize the activity with regard to safety aspects and authorize as well as supervise it. At the same time the state has to establish a legal basis for indemnification against the private actor which caused the damage, because international liability according to Art. VII of the Outer Space Treaty is only a liability in addition to the direct liability of the person who caused the damage.

The linkage of international liability to participation at the moment of launch of the space object implies that there is always a state liable toward the person damaged. The state from whose territory a satellite was launched is still liable for damage caused by that satellite several years later. The state which operated a satellite for several years is still liable if it hands over the satellite to an operator under foreign jurisdiction, after which the satellite causes damage. This seems to be unfair at first glance. But **states can agree among themselves** on an internal indemnification [9.4.15], [9.4.16]. If, in the latter case, the satellite was not operated by the state, but by a private entity, the state needs to receive information about the intention of that operator to sell the satellite to another operator abroad so that it can then enter into such an agreement with the state which exercises jurisdiction over the new operator.

This international liability established by Art. VII of the Outer Space Treaty is extended and specified by the Convention on International Liability for Damages Caused by Space Objects, as opened for signature on March 29, 1972 (**Liability Convention**).

Absolute liability as stated in Art. VII of the Outer Space Treaty is partly attenuated by Art. II of the Liability Convention. Only damages which are caused on the Earth's surface or to aircraft in flight (i.e., damages which might occur during launch or reentry)

are subject to absolute liability. For damages caused to other space objects or to persons or property thereon (i.e., damages in outer space) the launching state is only liable if it caused this damage, see Art. III Liability Convention.

Art. XI (2) of the Liability Convention clarifies that international liability of the launching state(s) according to Arts. II–IV of the Liability Convention pertains in addition to the "normal" claim for damages of the injured party against the injuring party. The injured party can choose whether it claims against the injuring party or against the launching state. Once the injured party has pursued one of these claims, the other claim is excluded.

Because the aforementioned rules are based on a treaty among states and because they intend to protect third parties, Art. VII of the Liability Convention excludes the applicability of this additional possibility to claim against the launching state(s) for damages to nationals of the launching state as well as for damages to persons who are participating in the operation of the space object which caused the damage. These persons may still pursue a claim for damages according to national laws or according to their contracts.

For a space object there is typically more than one launching state. This is based on the linkage of the participating of states with the launch event – as specified in Art. VII of the Outer Space Treaty and Art. I (c) of the Liability Convention. A state is liable if it launches the space object, if it procures the launching of the space object, or if the space object is launched from its territory or facility. If any of these criteria apply, the state is a launching state. For a German satellite launched, for example, from French Guiana (which is a French overseas department and therefore French territory) both France and Germany are launching states. The number of launching states increases correspondingly in international cooperation projects. The Liability Convention reflects this in Art. V: all launching states are jointly and severally liable. A state to which or to whose nationals damage was caused can address one of the launching states and claim the full damage from that state. This launching state must then internally claim compensation from the other launching state(s). The Liability Convention recommends that agreements among the launching states be concluded on how they apportion liability among themselves, Art. V (2). In case the space object is part of a private activity in outer space, the state(s) which might become liable for damage caused by that object need to know about such activity in order to decide whether or not to enter into such agreements with other launching state(s).

A claim for damages according to Art. VII of the Outer Space Treaty or Arts. II–IV of the Liability Convention can only be pursued by a state. If the damage was suffered by a private natural or legal person, they have to ask "their" state to place a claim against the launching state(s). The claim for damage has to be presented to a launching state through diplomatic channels within one year after occurrence of the damage (or identification of the launching state). If no settlement of the claim is reached within one year, the parties concerned must establish a Claims Commission composed of a member of the claimant state, a member of the launching state (if there is more than one launching state, they have to agree upon one) and a member chosen jointly by both parties. The decision of the Claims Commission is by majority vote. The decisions are final and binding.

Further Outer Space Law of the United Nations

So far, three main treaties on outer space have been mentioned: the Outer Space Treaty, the Registration Convention and the Liability Convention.

The Agreement on the Rescue of Astronauts, the Return of Astronauts and the Return of Objects Launched into Outer Space, as opened for signature on April 22, 1968 (the **Rescue Agreement**), has been of very little relevance so far. It obliges its member states to notify the launching authority as well as the Secretary General of the United Nations about every accident, condition of distress, emergency or unintended landing of a spacecraft in the territory under its jurisdiction. It specifies that the state takes all possible steps to render all necessary assistance to the astronauts and return them to the launching authority. Space objects returned to Earth have to be recovered and returned to the launching authority, as well. If the state on whose territory such an object was found has reason to believe that the object is of a hazardous or deleterious nature, it must notify the launching authority. The launching authority then has to take effective steps to eliminate possible danger.

The fifth international treaty on outer space is of almost no practical relevance. While the four aforementioned treaties have been ratified by many states (and are therefore binding for many states), only 2 of the 11 states which have ratified the Agreement Governing the Activities of States on the Moon and Other Celestial Bodies, as opened for signature on December 18, 1979 (the **Moon Treaty**), are spacefaring nations. A reason for the reluctance of states to ratify the Moon Treaty might be Art. 11 (1) of that treaty, which declares the Moon and other celestial bodies to be the "common heritage of mankind." The relevance of this norm, say for the use of mineral resources, is very much contested [9.4.17]. The (11) ratifying states have in any case agreed to establish an international regime governing the exploitation of mineral resources, see Art. 11 (5) Moon Treaty. Scientific investigation of the Moon is not restricted, see Art. 6 (1) of the Moon Treaty.

In addition to these five treaties, the United Nations has established seven **resolutions** on outer space. Resolutions are not international treaties. They do not bind the states. But resolutions are acts of volition of the (General Assembly of the) United Nations. They have the nature of guidelines [9.4.18]. The existing resolutions are as follows:

- A declaration of legal principles governing the activities of states in the exploration and use of outer space, adopted on December 13, 1963, which was substituted three years later by the Outer Space Treaty.
- The principles governing the use by states of artificial Earth satellites for international **direct television broadcasting**, adopted on December 10, 1982, which allows for direct broadcasting in foreign states (within the satellite's footprint) without the prior consent of that foreign state.
- The principles relating to **remote sensing** of the Earth from outer space, adopted on December 3, 1986, which allow remote sensing activities without the prior consent of the sensed state (while granting the right of access to the primary and processed remotely sensed data of its territory on a nondiscriminatory basis and at reasonable cost).
- The principles relevant to the use of **nuclear power sources** in outer space, adopted on December 14, 1992, which partly admit the necessity for using nuclear power sources in outer space and establishing safety requirements for them.
- A declaration on international cooperation in the exploration and use of outer space for the benefit and in the interest of all states, taking into particular account the needs of developing countries, adopted on December 13, 1996 [9.4.19].
- A resolution on applying the concept of the launching state, adopted on December 10, 2004, which provides recommendations in the context of private space activities [9.4.20].
- Recommendations on motivating states and international intergovernmental organizations to register space objects, adopted on December 17, 2007.

Some of these resolutions will be mentioned in the course of the following text.

9.4.1.2 Other International Outer Space Law

It is not just the United Nations that has originated treaties on outer space. Many other international and intergovernmental organizations have compiled treaties on or relating to outer space.

In 1961 the United Nations asked the International Telecommunications Union (ITU) to regulate radio communication with and through satellites. To this end, the ITU Convention and the Radio Regulations were amended, satellite services were defined and frequency bands were assigned [9.4.21], [9.4.22]. As the procedure for transferring frequency and orbit usage rights from the ITU to its member states is closely connected to the transfer of these usage rights from a member state to the satellite operator, this procedure will be dealt with under Section 9.4.2.1.

The World Trade Organization also regulated the area of telecommunication services. The General Agreement on Trade in Services including its annex on the liberalization of trade in telecommunication services entered into force in 1997. It also includes the opening of markets for satellite telecommunication services.

The International Civil Aviation Organization and the International Maritime Organization act in the field of satellite navigation. The World Meteorological Organization deals with regulations on access to the data of meteorological satellites.

Overall issues on intellectual property rights in outer space are discussed by the World Intellectual Property Organization. The International Atomic Energy Agency issues standards for the use of nuclear power sources on-board satellites; the International Organization for Standardization and the European Cooperation on Space Standardization develop standards for space projects and space technology. The UNIDROIT Convention on International Interests in Mobile Equipment and its draft protocol on space assets raise issues on the simplification of financing for space projects [9.4.23], [9.4.24].

The United Nations Educational, Scientific and Cultural Organization has developed rules on ethics in outer space.

Intergovernmental agreements exist on the development and use of the International Space Station [9.4.25], on the operation of satellite communication systems (ITSO and IMSO) and on satellite meteorology (EUMETSAT and ARABSAT).

Space activities often have military implications. Some aspects are directly dealt with in the aforementioned treaties (e.g., the peaceful use of the Moon and other celestial bodies or the prohibition to place in Earth orbit objects carrying nuclear weapons or weapons of mass destruction). In addition, military treaties of indirect relevance to space activities have to be taken into consideration (e.g., the Anti-Ballistic Missile Treaty or the Nuclear Test Ban Treaty or the Comprehensive Nuclear Test Ban Treaty). Finally the (voluntary) Missile Technology Control Regime has to be mentioned [9.4.26].

Lastly, many bilateral and multilateral agreements are in force covering general scientific–technical cooperation, special outer space applications and specific space missions [9.4.27].

9.4.2 Legal Conception of a Space Mission

Even when planning a space mission many legal issues have to be considered. What licenses might be necessary for which activities and how they might be obtained will be described in the following, using Germany as a specific example. When and to what extent space objects have to be registered will also be described. Finally, some possible contractual relations as well as liability scenarios are examined. This section is intended as a checklist including practical guidelines, without a claim to completeness.

9.4.2.1 Necessary Licenses

Frequency Assignment

An operator needs frequencies within certain frequency bands (e.g., C- or Ku-band) in order to command a satellite as well as to transmit data between the satellite and its ground segments. Usable frequencies are a limited resource for technical reasons. Therefore, there is a need to **coordinate** the use of frequencies.

Such coordination is two tiered. First, the use of frequencies by the ground station has to be coordinated and assigned on a national level. Second, the use of frequencies in outer space has to be coordinated and assigned internationally by the ITU for its member states; the member state can then assign the right to use such frequencies to an operator. This international coordination and assignment is dealt with in the following section.

The use of frequencies by ground stations requires prior allocation by the Federal Network Agency. The procedure of allocation is laid down in §55 of the Telecommunications Act.

Simple reception of telecommunication satellite signals as well as the use of very small satellite senders and receivers are typically approved by a general license. They can be set up and used without prior allocation of frequencies. But a specific license is necessary to use frequencies in order to command a satellite as well as to transmit data from a satellite to a ground station.

The person intending to use such frequencies has to apply at the Federal Network Agency. The application should specify the service and type of network, the technology and the area in which the frequency will be used. The applicant has to demonstrate the ability to use the frequencies effectively and without interfering with other allocated frequencies. Application forms can be downloaded from the website of the Federal Network Agency [9.4.28].

The allocation of satellite frequencies for ground segments is approved if the frequencies are identified

in the frequency usage plan [9.4.29] and available. The use of such frequencies needs to be efficient, without technical disturbance, and compatible with other frequency uses.

The applicant has to pay a fee. According to the Frequency Fee Ordinance, the Federal Network Agency (currently) charges between €68 and €1000. Besides these fees, the applicant has to pay an annual fee of (currently) roughly €25 per radio station according to the Ordinance on Contributions to Protect Interference-free Frequency Use.

The licensee must indicate the starting and termination dates of the frequency use to the Federal Network Agency. The licensee also must report subsequent changes in personal circumstances, such as changes of name, address or shareholder structure.

As ground segments are often operated by a party other than the operator of the satellite, the person to whom frequencies are allocated for the ground segment does not need to be the same person to whom frequencies are allocated for use in outer space. Cooperation between the ground and space segments is regulated by a treaty among the parties, see Section 9.4.2.3.

Transfer of Orbit and Frequency Usage Rights

While the allocation of frequencies according to §55 of the Telecommunications Act regulates the use of frequencies by the ground segment, the use of frequencies in outer space (e.g., to command a satellite) requires additional approval. The operator of the satellite has to apply at the Federal Network Agency for a transfer of orbit and frequency usage rights. These usage rights are allocated by the International Telecommunications Union (ITU) to the member states (in this case, Germany). This approval is necessary in addition to the allocation of frequencies for the ground station, because the use of frequencies by satellites needs international coordination, which is handled according to the public international rules of the ITU [9.4.30].

The procedure for the **transfer of orbit and frequency usage rights** is regulated in §56 of the Telecommunications Act and Administrative Order 8/2005 of the Federal Network Agency [9.4.31].

At a very early stage the German operator of a satellite system notifies the Federal Network Agency of its intention to use satellite frequencies. If the international registration of the satellite system is not via the Federal Network Agency, the operator must provide evidence to the Federal Network Agency of the initiation of the international procedure by another ITU member state.

If international registration is via the Federal Network Agency, the operator must declare one year before launch its intention to commission the satellite system and describe the technology to be employed and the operating concepts for the satellite.

The Federal Agency then applies for the needed obit and frequency at the ITU with the goal of having the orbit and frequency usage rights transferred to the Federal Republic of Germany. The Federal Network Agency can permit the private operator to exercise these rights only after these orbit and frequency usage rights have been transferred to Germany; that is, after completion of the ITU procedure.

According to the procedures of Administrative Order 8/2005, the private operator has to apply at the Federal Network Agency so that the Federal Network Agency can apply for these rights at the ITU. The application of the operator has to contain information about the applicant, explanations of how international coordination is to be assured, acceptance of the obligation to comply with ITU regulations, confirmation of a fiduciary account or a bank guarantee in the amount of the expected ITU fees, and a concept for the use of the frequencies. The basic technical parameters of the satellite systems have to be supplied as well, especially the frequency band and orbital position to be used, as well as the intended footprint.

The application is published as advanced publication information by the ITU. This is intended to provide information about the planned use of the orbit and frequency to all other ITU member states. Based on this information, other member states can assess whether there might be interference with the frequencies they use. At least six months after the advanced publication information, ITU issues the coordination request. It includes a very detailed technical description of the satellite system. In case the applied frequency might interfere with (existing or planned) other frequency usages, the member state concerned can (also on behalf of its nationals) object to granting the requested frequency by filing an appeal with the ITU.

The ITU collects and forwards to the Federal Network Agency these objections to the frequency application. The Agency is now called on to **coordinate** the frequencies with those ITU member states which have filed an appeal. The responsibility for this coordination is transferred by the Agency to the satellite operator who has applied for the frequency since the outcome of coordination has a considerable effect on the detailed system design, the operational running of the satellite network and economic aspects. Nevertheless, the Agency assists the applicant in effecting international coordination insofar as the applicant lacks entitlement to carry out the ITU procedures, given the private law status of the applicant. In particular, the Agency forwards any document prepared by the applicant to the ITU or other member states "on behalf of Germany." The Agency assesses these documents for conformity with ITU rules and regulations as well as with regard to the public interests involved.

In practice, the international coordination lasts several months (sometimes even years), depending on the rights which have to be coordinated. In addition, the applicant has to be aware that the international coordination might necessitate changes in the satellite system design in order to achieve compatibility with other satellite systems. This calls for a flexible mission plan.

As soon as the international coordination is finalized, the Agency notifies the results to the ITU, supplying the final satellite system design. The Agency is assisted by the applicant.

The ITU finally enters the satellite systems (especially the telecommunication links) in the Master International Frequency Register. From then on, the frequency usage is protected by international law and must be respected by all ITU member states (including their nationals). The Federal Republic of Germany receives the right to use the orbit and the frequency.

The German satellite operator may now apply at the Agency, which holds this right of usage on behalf of Germany, for a transfer of these usage rights. The application must be filed at least three months in advance of the scheduled launch of the satellite system. The transfer of the right to use the orbit and the frequency is a frequency assignment, comparable to – but not identical with – the frequency assignment for ground segments according to §55 of the Telecommunications Act.

The rights to use the orbit and the frequency will be transferred to the applicant if the frequencies are available, compatible with other frequencies used or filed, and do not contravene public interest. The applicant has to prove that indeed both the orbit as well as the frequency are used. The applicant also has to prove that the transmission does not cause interference, that there is operational control over the space segment, and that competent personnel are employed. The applicant finally has to assure proper operation of the satellite system.

If the launch of the satellite system into orbit is postponed, the applicant has to inform the Agency as soon as possible. Right after launch the applicant has to inform the Agency about the launch as well as the final orbital parameters. The applicant must also register the space segment as a space object.

The transfer of the rights to use the orbit and frequency is limited to the expected lifetime of the satellite system. The tables of transmitter frequency tolerances and maximum permitted spurious emission power levels have to be followed. In case of interference with other satellite systems or interference with other transmissions, the Agency may restrict or prohibit the use of the frequency. It also may revoke the transfer of rights to use this orbit and frequency.

The operation of the satellite for which orbit and frequency rights were transferred must be planned so that there are enough reserves (of energy, fuel and functionality) to carry out decommissioning orbital maneuvers in accordance with international standards. Operating requirements may be imposed on the party enjoying the right of use in order to achieve compliance with international standards on the avoidance of space debris.

There are charges for both the ITU application and the transfer of the right of use. In addition to the fees of (currently) €500 to €3500 according to the Frequency Fee Ordinance of the Agency, the applicant also has to pay the expenses for the ITU procedures (currently 570 Swiss francs for the advanced publication, 5560 to 33 467 Swiss francs for the coordination and 15 910 to 57 920 Swiss francs for the notification).

Export Control License for Space Technology

Private industry (as well as research institutions and universities) exporting goods or technology must nominate a person responsible for all export activities. This person has to be a member of the organization's management or executive board. The person is responsible for assuring that the export control regulations can be met. To this end the person must select and supervise appropriate personnel and ensure that they are adequately trained. The person responsible is personally liable for all violations of export control regulations.

It is forbidden to export **weapons of war**, to export economic resources (e.g., goods, money, services, know-how) to people, organizations or other entities listed on the **EU terrorism list,** as well as to export into states which are under **embargo**.

Weapons, ammunitions or armaments which are named in the export list (Part I, Section A) and which are intended to be brought into another country (including a member state of the EU) need to be licensed for export control. This is usually not relevant for space industries, research centers and universities.

It is more relevant that **dual-use technology** also needs to be licensed. Many space goods and space technologies are dual-use items. Dual-use technology refers to all goods (including data processing programs as well as technology) which may be used for both military and civil purposes. Because there are no launch sites on German territory, space goods developed or assembled in Germany typically have to be brought outside of Germany for launching them into outer space. In order to know whether or not a specific good is deemed to be a dual-use item, the export list (Part I, Section C) has to be studied. This list provides an exhaustive enumeration of all dual-use items. It includes communication items (Category 5 of the list), optical and radar sensors (Category 6 of the list), propulsion devices, rockets, specific parts of rockets, sounding rockets and reusable launch vehicles (Category 9 of the list). For each of these categories the list identifies technical specifications which have to be fulfilled in order to be subject to the export control regime [9.4.32]. If these specifications are not met, the goods can be exported without a license. The export list can be downloaded from the Internet [9.4.33].

The export list contains not only goods (to be found in Category A of each section) but also technology (to be found in Category E of each section). Technology is understood to be specific technical knowledge which is necessary for the development, production or use of goods and products. This technical knowledge has to be included in documents, namely blueprints, schemes, diagrams, models, formulas, tables, construction plans or specifications, descriptions or instructions. It has to be in writing or otherwise recorded, for example on magnetic disk or tape. Technical knowledge can also be exported in the form of technical assistance, namely training, schooling, assisting or advising. Technical assistance can also be understood as the dissemination of technical documents. Of practical importance are efforts of foreign nationals (non-Europeans) to establish ties at congresses or trade fairs. Technical knowledge may also be exported in research and development projects with non-European industries, research centers or universities as well as in training and schooling programs. It may also be exported in the exchange of researchers, engineers, doctoral students and students. Technical documents may also be transferred in order to enter into insurance contracts. In contrast, if technical documents are transferred in the context of a procurement tender, the bidder is exempted from applying for an export control license.

The need to apply for an export control license for dual-use items depends only on whether or not the good is listed – irrespective of which state the good is going to be exported to. Only goods brought into a member state of the EU are not subject to an export control license. In exceptional cases there is a need to apply for an export license for goods which are not listed, if the person intending to export the good has indications that it will have a **military end use**, for ABC weapons, for missiles for such weapons or for nuclear purposes.

Persons intending to export a listed good are asked to send a preliminary application to the **Federal Office of Economic and Export Control** (BAFA). This serves to clarify whether there is a need for an export control license or whether such a license can be issued at all.

The license has to be applied for at the BAFA. The legal basis is **Regulation (EC)1334/2000** of the

Council of June 22, 2000 which set up a Community regime for the control of exports of dual-use items and technology. As an EC Regulation, these rules are directly in force in every member state as national law. In addition, further restrictions exist on the national level for dual-use items based on §§7 *et seq.* of the Foreign Trade and Payments Act.

Application forms can be downloaded from the Internet [9.4.34]. The applicant has to include technical documents, for example brochures and data sheets, to provide a basis for technical evaluation of the good or technology. In general the applicant should also include an official or private end use certificate. The BAFA examines (if necessary in cooperation with the Federal Ministry of Economics and Technology and the Federal Ministry of Foreign Affairs) on a case-by-case basis whether the export endangers **national security or foreign policy interests**.

Special conditions and procedures exists for exports of goods into a state against which an embargo is in force (presently, Armenia, Azerbaijan, Bosnia, Haiti, Iraq, Libya, Myanmar, Rwanda, Serbia/Montenegro, Sierra Leone, Somalia, Sudan, Zimbabwe) or which is listed on Country List K (Cuba, Iran, Lebanon, Mozambique, North Korea, Syria).

The procedure for granting a license takes from two weeks up to three months. In difficult cases, especially if an embargo or a state listed in Country List K is involved, the procedure may last up to 12 months.

Infringement of export control regulations may be subject to payment of a fine of up to €500 000 or imprisonment up to five years, in severe cases up to 10 years.

US export control regulations are of factual importance in the space sector in addition to German and European export control regulations. In many domains US industry is dominant; some components can only be built in the USA.

The importance of US export control regulations is only indirect insofar as a German entity might need an export control license from the US State Department in order to export the goods from the USA to Germany. But in addition, the State Department grants this license only subject to conditions with respect to the re-export of the good (including related knowledge). This means that the State Department sets conditions if the good is to be exported from Germany into a third state. This is the case, for example, if a component exported from the USA to Germany is being integrated into a satellite in Germany and this satellite is then sent to a launch site, for example, in French Guiana or Kazakhstan. Although the re-export of US goods without permission by the USA cannot be a reason for infringement of German export control regulations, there is an indirect consequence. By non-compliance indirect (economic) pressure may be put on the German entity regarding future cooperation with US industry. This is why the German recipient of goods imported from the USA should always indicate into which states it intends to re-export the goods (even after integration into a satellite) [9.4.35], [9.4.36].

License to Launch and Operate a Space Object

Those states which have ratified the Outer Space Treaty have committed themselves to establishing national procedures to authorize and supervise non-governmental activities in outer space, Art. VI (2) Outer Space Treaty. In any case, the operation and control of a satellite, a probe, a platform or a space station is an activity in outer space. Considering the practice of states, also such activities have to be understood as being undertaken "in" outer space if they are only partly undertaken in outer space but directed toward exploration or use of outer space. This includes launching activities. But according to the opinion of most scholars, the launching of sounding rockets is not included, nor are suborbital launches. An activity is deemed to be "non-governmental" if it is undertaken by a private company, a research center or a university.

Consequently, a state has to implement a **procedure to authorize** and supervise activities in outer space, when, for example, the first satellite operated by a private company or university is going to be launched into outer space. Such a procedure may be established in the context of already existing administrative procedures (e.g., as a condition to the decision to give a grant to the project of the operator). It can also be part of a contractual relationship between this operator and the government, which might be the case, for example, if the government is commissioning the task to operate the satellite to a private company (e.g., in the context of a public–private partnership).

Anyhow, it seems to be more appropriate to establish such a procedure by separate legislation, which might be called **national space legislation**. This would be the most transparent approach for the private operator, research center or university and provide the most legal security.

There is no legislation in force in Germany which can deal with the authorization of space activities. But some private companies and universities already exist which operate satellites in outer space, and also a German launch service provider launches from a launch site abroad, so that the German government needs to ensure authorization of these activities by means other than legislation. Analyzing these other means, one cannot identify a coherent procedure.

Taking into account the international responsibilities Germany assumed as any other state by ratifying the Outer Space Treaty, and taking into account as well the existing legislation abroad, certain aspects can be elaborated for the contents of an authorization procedure.

First, it has to be ensured that the private entity, research center or university carries out its activities in outer space in conformity with those rules which are binding for the state ratifying the Outer Space Treaty. In this context, the state also has to ensure the safety of such activities in order to avoid damage and liability. All this can be done within the **authorization**.

The obligation to ensure the conformity of non-governmental activities with the activities of governments or governmental agencies is explicitly mentioned in the Outer Space Treaty, see Art. VI (1). The state is responsible for ensuring that non-governmental activities adhere to the principle set forth in the Outer Space Treaty. They have to follow the same obligations which have been dealt with under Section 9.4.1.1.

In the first instance, the obligation to ensure safety of an activity in order to avoid damage and liability has a regulatory as well as a social character. But Germany also has a strong self-interest in avoiding damage in the course of an activity in outer space. This is because Germany is liable as a launching state for damage caused by a space object to a third party according to Art. VII of the Outer Space Treaty and the Liability Convention. This liability also exists if Germany was not involved in the operation which finally caused the damage (e.g., collision of a privately operated satellite with another satellite due to a lack of fuel at the end of its operational life), because international liability is assumed by the involvement of the state at the moment of launch (i.e., from the very beginning of a mission) of the object into outer space. Full international liability exists, in addition to the direct liability of the person who caused the damage to the person damaged, according to national tort law. This was dealt with under Section 9.4.1.1. Therefore, it is part of the authorization procedure for the operator to demonstrate its **reliability**, **capability** and **technical qualification** also in order to avoid such damage. In addition, the state ensures the **safety** of the products used (e.g., on the basis of technical standards such as ISO, DIN or ECSS). Finally, compulsory **third-party liability insurance** is required.

Besides implementing its international responsibility, according to Art. VI of the Outer Space Treaty, and besides ensuring the avoidance of liability, a state has a third aspect of importance to consider. This is also out of self-interest to some extent. Based on an application for authorization, a state can assess whether or not it is the only responsible state for the activity to be authorized, or whether other states may be liable for the object (satellite) which is used in such activity. If it is also a launching state, it receives in the application the information necessary to enter into negotiations with other launching states regarding the internal apportioning of damages to be paid, see Section 9.4.1.1.

Germany not only has to authorize an activity at the beginning of a mission, but also has to supervise it throughout its entire duration. On the basis of legislation or by other means the state has to obligate the operator to inform the responsible authority on a regular basis or in case of incidents. The state also has to ensure that it always can demand additional information as well as that it has the right to enter the premises of the operator to make inspections. The operator might object if this is done without adequate legal basis. Finally, only on the basis of a formal legal act can the state assess fines or other punishment in case of nonapplication for an authorization or infringement of the terms and conditions of the authorization or the legislation.

A special aspect of **supervision** is the need for the authorizing authority to receive information in case the satellite (with which the activity in outer space is undertaken) is to be purchased by a foreign national. This is because the state, being also the (liable) launching state, will remain the liable launching state also after the purchase, as the liability is linked to the moment when the launch took place. If the new (foreign) operator causes damage, Germany remains liable, although Germany cannot supervise the foreign operator. Therefore, the launching state has to enter into negotiations with the state which is responsible for the activity for which the satellite was purchased. That foreign state should agree to assume liability *inter partes*.

Operators of a satellite (or the person who wants to undertake a space activity) should take a close look at not only the national legislation, but also the foreign authorization procedures. This is especially important in the context of increasing international cooperation, also because a German operator always has to launch its satellite from abroad. Even if the German operator does not apply for the foreign authorization, the mission concept can be influenced by the regulations which the foreign partners and contractors have to follow.

Formal national legislation with regard to outer space activities exists as of 12 states [9.4.37]:

- USA (49 USC 701 Commercial Space Launch Activities; CFR 14 III 400 Commercial Space Transportation)
- Norway (Act on Launching Objects from Norwegian Territory etc. into Outer Space, 1969)
- Sweden (Act on Space Activities, 1982; Decree on Space Activities, 1982)
- United Kingdom (Outer Space Act, 1986)
- South Africa (Space Affairs Act, 1993 as amended)
- Russia (Law about Space Activity, 1996 as amended; Statute No. 403 Licensing Space Operations, 2006)
- Ukraine (Ordinance on Space Activity, 1996 as amended; Decree No. 798 on Licensing Private Entities undertaking Space Activities)
- Australia (Space Activities Act, 1998 as amended, Space Activities Regulations, 2001)
- Hong Kong (Outer Space Ordinance, 1999)
- Brazil (Administrative Edict No. 27, 2001)
- Belgium (Law on the Activities of Launching, Flight Operations or Guidance of Space Objects, 2005)
- South Korea (Space Exploration Promotion Act, 2005).

It is know that the following states are drafting or plan to draft similar legislation: the Netherlands, France, Italy, Luxembourg, Indonesia, India, Malaysia and Kazakhstan.

Authorization and supervision of non-governmental activities is often the motivation as well as the essential reason for enacting national space legislation. But most states have connected the legal basis for a registration system with specific liability issues, especially rules on indemnifying the state which has been held liable as a launching state according to Art. VII of the Outer Space Treaty and the Liability Convention. These aspects will be dealt with separately in Sections 9.4.2.2 and 9.4.2.4.

Authorization of Space Vehicles as Aircraft

As long as there is no national legislation on authorizing space activities in force in Germany, operators should always take aviation legislation into consideration. This is because the legislator has rudimentarily provided for the applicability of this legislation to space activities: §1 para. 2 Air Traffic Act states that space vehicles, rockets and similar objects are considered as aircraft, as long as they are in space.

This would be the basis for applying regulations of the Air Traffic Act, the Air Traffic Regulation and the Air Traffic Approval Regulation also to space vehicles.

But the fiction established by §1 para. 2 of the Air Traffic Act only applies as long as the space vehicle, rocket or similar object is in German airspace. This airspace ends at an altitude of something in between 83 and 100 km, depending on which expert opinion one follows. Some experts argue that airspace ends at the so-called Kárman line, which is where the airspace can still slightly support aircraft and where orbits of space vehicles are also possible (physical approach). Others argue for a regulatory limit of 100 km. But a space vehicle is very unlikely to cross German airspace, as there is no launch site either in Germany or in its neighboring countries. Hence, the fiction is almost without any practical applicability.

In case there should one day be an activity of a space vehicle in German airspace, the fiction would apply. Nevertheless, many regulations in the Air Traffic Acts are not apt to apply to space activities. A space vehicle does not need to be certified according to §2 of the Air Traffic Act in combination with §1 Air Traffic Approval Regulation. But the operator might need a flight crew license according to §4 Air Traffic Act, an approval for ascent according to §16 Air Traffic Order, as well as an approval to leave German airspace according to §2 para. 4 Air Traffic Act.

Notification of Telecommunication Networks

As soon as public **telecommunication services** are provided via satellites, notification is required. This is in addition to notification regarding the frequency assignment and the transfer of rights to use the orbit and frequency. Notification is made to the Federal Network Agency in accordance with §6 of the Telecommunication Act.

The notification has to include information which can identify the operator or service provider. This contains, for example, its number in the commercial register, physical address, a short description of the network or service, as well as the estimated date of the commencement of service.

Authorization to Operate an Advanced Remote Sensing Satellite System

On December 1, 2007 the German Act to safeguard security interests in the distribution of remote sensing satellite data (Satellite Data Security Act) came into force [9.4.38]. According to this Act the **operator of an advanced remote sensing satellite system** has to apply for authorization. A remote sensing satellite system is deemed to be "advanced" if the system is able to generate data of high information content. This information content is determined according to the spatial **resolution**, spectral coverage, number of spectral channels, etc. The exact technical limits are determined in the Remote Sensing Satellite Regulation [9.4.39].

The operator of such a remote sensing system has to apply for authorization at the Federal Office of Economics and Export Control (BAFA) prior to the launch of the system. The BAFA evaluates whether the operation of the remote sensing satellite system might endanger German **security or foreign policy interests**. Therefore, the BAFA checks the reliability of the operator as well as the operational security. The operator has to ensure that the satellite system cannot be commanded and data cannot be accessed by unauthorized persons. Furthermore, the operator has to ensure that no unauthorized person has admission to the rooms from which the satellite system is operated and which are used to store and handle any associated data, and that no unauthorized person has access to relevant facilities therein. Persons having such access must have a security clearance.

The BAFA is supported by the Federal Office of Information Security.

The Act will be supplemented by an ordinance on fees, which has not yet been published.

Permission to Act as a Data Distributor

The Satellite Data Security Act also requires a person who intends to distribute advanced remote sensing satellite data to apply for permission. This permission is granted by the BAFA.

A **data distributor** is anybody who circulates such data for the very first time to customers. The operator of a satellite system can also be a data distributor if it circulates data to persons who have not themselves received permission to act as a data distributor. Usually the operator will transfer the right to use the satellite data to a data distributor, who in turn will circulate the data to customers. In this case, only the distributor needs to apply for permission.

The BAFA evaluates (similar to the evaluations to authorize an advanced remote sensing satellite system) whether the distribution might endanger German **security and foreign policy interests** and whether the data can be transferred on a secure basis. This includes ascertaining the reliability of the data distributor. The distributor also has to ensure that no unauthorized person has admission to the rooms relevant to the storage, handling and distribution of the data and that no unauthorized person has access to relevant facilities therein. Persons having such access must have a security clearance.

A person who intends to distribute data is not obliged to apply for registration as a data distributor. In that case the person is dealt with as a (simple) customer. This has the consequence that the person

only receives data after the operator has checked the data according to how sensitive it is (and if sensitive, that distribution has been licensed by the BAFA). This will be described in the following section.

License to Distribute Remote Sensing Data with High Informational Content

Besides the authorization to operate an advanced remote sensing satellite system and besides the approval to act as a data distributor, the core regulation of the Satellite Data Security Act requires the distributor to apply for a **license to distribute remote sensing data** with high informational content (if the request for data is sensitive). This license is necessary for every distributor who circulates data for the very first time to customers.

On the first level, the data distributor who intends to distribute data to any persons who are not themselves registered as a data distributor has to perform a **sensitivity check**. This sensitivity check is related to the request for data (and not related only to the contents of the data). The sensitivity check assesses whether the intended distribution of the data might endanger German **security and foreign policy interests**. This assessment is made according to the following criteria:

- Technical parameters of the data.
- The sensed territory.
- The time of generation of the data, the time of delivery and the person to be supplied.
- The ground segments to which the data is to be sent.

The assessment should consider the person who requested the data as well as any persons who are intended to have access to the data. The assessment is made according to criteria specified in the Satellite Data Security Regulation. These criteria can be automatically assessed by the data distributor.

The data distributor can continue the transaction if the sensitivity check reveals no endangerment of German security or foreign policy interests. Not until the check reveals sensitivity does the distributor have to decide whether to decline the transaction or apply for a **license** from the **BAFA**. Here an administrative procedure starts. The BAFA will ascertain whether the transaction might indeed endanger national security or foreign policy interests.

The license procedure is supposed to be rapid, and if possible concluded within a month of the date of application. Many applications are decided within a much shorter period. The only criteria against which the application can be assessed concern whether the transaction of the data would indeed endanger German security or foreign policy interests.

Similar legislation exits in the USA [9.4.40] and Canada [9.4.41], [9.4.42]. Only these two states currently have advanced remote sensing satellite systems not solely used for military purposes.

Particular Case: Grants for Space Missions

Not only licenses but also **grants** establish a legal relationship to the state. Grants are subsidies which are given within the framework of the national space program [9.4.43] for space missions which are in the national interest and which would not be possible without a state subsidy. The legal basis for grants is to be found in §§23, 44 of the Federal Budget Act.

Grants are usually approved for projects and are subject to reimbursement. If the mission is within the domain of basic research it can be subsidized fully; if it is an industry mission it can be subsidized up to 50%. Precompetitional developments may be subsidized up to 25%. The European Commission allows supplements for small and medium-sized enterprises as well as for cooperation projects.

Anybody interested in applying for a grant may contact the German Space Agency within the **German Aerospace Center** for advice on the feasibility and prospects of an application. They may then apply for a grant. The application should include among other things a description of the project, a financial budget and a preliminary costing.

The Space Agency will decide on the application on behalf of the Federal Ministry of Economics and Technology. It assesses the creditworthiness of the applicant with respect to the costs covered by the applicant, the budget in the context of grants, the financing plan and the feasibility of the project. It should be borne in mind that a grant is excluded if the applicant has already started the project.

The project is monitored by the Space Agency while it is being carried out by comparing the progress to the original work, time and budget planning. Forward projections as to realization are monitored as well.

The Space Agency checks whether the grants are used appropriately. Accordingly, the grantee has to provide supporting documents on a regular basis as well as final documentation on how the grants were used in the project. The Space Agency determines whether the intended purpose of the grant was fulfilled.

The decision to give a grant is supplemented by a fixed set of terms and conditions which are not subject to change. For grants based on expenditures, the grantee has to follow the general terms and conditions for grants subsidizing a project (ANBest-P) as well as some specific terms and conditions (BNBest-BMBF 98). For grants based on costs, specific terms and conditions exist (NKBF 98). The general terms and conditions mandate economical and cost-effective use of the grant, and how it is to be spent. For grants exceeding €100 000, tenders must be sought for all contracts that the grantee intends to conclude in the course of the project, as if the grantee were a public entity. One of the most important terms within the specific terms and conditions is that any intellectual property which might be generated in the course of the project be owned by the grantee. Ownership includes the exclusive right of use, but the grantee is obliged to actually use this intellectual property, since it was generated in a project which was financed to some extent with public money, in accord with an exploitation plan drawn up in cooperation with the Space Agency. The intellectual property must be used for innovation, and the grantee is obliged to file an application for a patent if necessary. Here, the grantee has to be aware that once the intellectual property has been published it is no longer a new invention (and thus not patentable).

9.4.2.2 Registration of Space Objects

Objects which are launched into Earth orbit or beyond have to be entered in a national registry. States have assumed the obligation to hold such a registry, see Section 9.4.1.1.

The Federal Republic of Germany has established such a registry. It is an annex to the Aircraft Registry which is maintained by the **German Aviation Authority** (Luftfahrt-Bundesamt). As of June 2008, 31 space objects were registered in the German registry.

Space objects owned by the government are registered on the initiative of the German Aviation Authority or on notice of the Federal Ministry for Foreign Affairs or the Space Agency within the German Aerospace Centre. Private space objects are so far registered by the voluntary notice of the owner or operator. An obligation to provide information for registration to the national aviation authority does not exist within German legislation. Aircraft are registered according to §3 of the Air Traffic Act in combination with §14 of the Air Traffic Approval Regulation. But neither space vehicles nor space objects are mentioned in §14 of the Air Traffic Approval Regulation. Also, the fiction established by §1 of the Air Traffic Act does not apply to these norms, according to the wording (see above).

The "obligation" specified by Administrative Order 8/2005 in the context of the transfer of orbit and frequency rights (see above) to give notice of a space object which is to be launched into orbit to the aviation authority can be regarded as a request for information only. This is because such an Administrative Order cannot constitute a formal legal basis for such an obligation.

Nevertheless, it is highly recommended to give notice of a space object to the aviation authority on a voluntary basis. This recommendation is based on the implications which registration has for several aspects (see Section 9.4.1.1).

The operator/owner should give notice of space objects for which Germany is considered to be a launching state. A space object is an object which is launched (or intended to be launched) for the exploration or use of outer space [9.4.44], [9.4.45]. As the objects need to be launched into an orbit (at least describing one full orbit), these typically include satellites, probes, platforms as well as rocket stages which remain in orbit for a while. On the other hand, sounding rockets and suborbital launch vehicles are not included.

Germany is a launching state also for privately owned/operated space objects if the object was launched from German territory or if Germany procured the launch to be carried out. As there are no launch facilities within Germany and as the German government does not own a launch site on a foreign territory, the criterion "to procure a launch" is of special importance. The interpretation of this notion was given in Section 9.4.1.1 above. Germany procures a launch if the German

government has authorized the activity which is undertaken or if Germany was obliged to authorize such activity. It might also be the case if Germany has otherwise contributed to launching the object; for example, by providing grants to the operator/owner, by concluding contracts for the manufacture and/or launch of the object with the owner/operator, or by having a scientific interest in the operation of the object. The notion "to procure" is to be interpreted very widely, as can be seen. If the object is operated by a German national or by a German private entity (company, university, research center), this is an indication that Germany also procures the launching of this object and therefore is a launching state.

The content of the German registry on space objects resembles the international register held by the (Secretary General of the) United Nations. Therefore, the operator should provide the following information to the German Aviation Authority: the foreign governments or foreign non-governmental entities involved in the launch and operation of the object (in order to determine the other launching states), an appropriate designation for the object, the date and location of the launch, the basic orbital parameters (nodal period, inclination, apogee, perigee) and the general function of the space object.

As the aviation authority only registers the object after the launch took place, simple notification of that fact is sufficient. Nevertheless, it is recommended that the aviation authority be informed in advance. This is because there will always be another launching state in addition to Germany (since the launch has to take place from a foreign territory). Consequently, the German Aviation Authority (ideally in cooperation with the Federal Ministry of Foreign Affairs) should enter into agreements with this foreign state. These **agreements with other launching states** include the determination of which launching state will register the object in its national registry, since only one of the launching states can do this (see Section 9.4.1.1). Only when this agreement has been signed, and only if it identifies Germany as the launching state which registers the object, can the German Aviation Authority register the object in the annex to the Aircraft Registry.

The German registry is not public. It can be accessed by applying to the aviation authority. The application has to state the reasons why the information is needed. Other foreign national registries on space objects are accessible, for example on the Internet [9.4.46], [9.4.47], [9.4.48], [9.4.49].

Right after an object is registered in the German registry, the aviation authority informs the Federal Ministry of Foreign Affairs about the registration. It provides the information contained in the registry to the Federal Ministry. The Ministry will notify the Secretary General of the United Nations about the registration. The Secretary General (supported by the Office for Outer Space Affairs) will enter this information in the International Register of Space Objects, see Section 9.4.1.1.

9.4.2.3 Contractual Conception of a Space Mission

To prepare and undertake a space mission requires the operator to enter into many contracts. The necessity to conclude a contract as well as the contents and conception of such a contract are specific to each particular case and depend on many factors, such as which parts of the outer space missions are performed by the operator. In the following section some exemplary contracts are described, assuming that the contracting party is the operator/owner of the space object (e.g., satellite). But as the contracting parties enjoy the general principle of freedom of contracts, only some general examples and recommendations can be given.

Manufacture of Space Objects

Specific definition of the **main contractual obligations** is the core part of a contract on developing and manufacturing a space object. With regard to the obligations of the manufacturer, this is usually done in a very extensive, technical description which is usually annexed to the contract. But the contracting parties may not ignore the fact that the description of the main contractual obligations is an essential and important part of the contract itself and cannot be totally allocated to an annex. The fundamental obligations have to be dealt with in the contract. It is only the description of the (technical) details which may be put into the annex. Standards exist for many

technical descriptions, including those for parts and appliances of space objects. These may be referenced in order to ensure high quality. DIN, CEN, ISO and ECSS standards are of special importance. Some other standards for space projects also exist in the context of AECMA. Such standards are not legal acts and therefore not applicable as such. They have to be agreed upon by the contracting parties. Many space agencies as well as many industries (being themselves represented in Eurospace, the association of European space industries have voluntarily committed themselves to use ECSS norms. Standards ensure not only good quality but also functional integrity and compatibility.

Whether or not the manufacturer has fulfilled its obligations has to be determined according to the descriptions set forth the contract. Note that the parties should agree whether the manufacturer is obliged to render a service to the purchaser, or deliver a result (e.g., hand over a product). This is because, having agreed on a service, the manufacturer has already fulfilled its obligations as soon as it has rendered the service and attempted a positive result. The parties will agree upon a service contract if the manufacturer is asked to develop a new technology (e.g., to develop a new type of satellite bus). On the other hand, if the parties agree on a result to be achieved, the manufacturer has to deliver exactly this result. It is no longer sufficient for the manufacturer to attempt to receive a result. The parties will agree upon such a contract if a well-established product is to be manufactured, for example a current type of satellite bus equipped with conventional instruments. It is important to be aware of which of the two types of contracts was agreed upon, as this determines the moment when the manufacturer has fulfilled its obligations; that is, the moment when the purchaser can no longer claim subsequent improvement or compensation for damage.

A manufacturer may also give a guarantee, going beyond the simple obligation to render a service or deliver a result. But by doing so the manufacturer has to be clear that this implies additional risks. If the manufacturer does not satisfy the guarantee, the manufacturer will be subject to claims for compensation, even if it was not responsible for not fulfilling its obligations.

The main obligation of the purchaser will be payment of a certain sum of money. The contract should deal not only with the amount to be paid, but also with the means of payment (e.g., fixed price or reimbursement of expenses) and the terms of payment (milestones depending on fixed dates, or performance related) [9.4.50]. The purchaser should monitor fulfillment of the contract in technical reviews.

The contracting parties should always stipulate the risk of delays. Space missions are typically long-lasting projects, which intensifies this risk. The parties should allocate the risk of delays in different phases of the contract to the manufacturer or the purchaser. The establishment of different phases makes it easier to amend the contract if necessary.

Contracts (especially on research and development) may specify **non-disclosure** of information and know-how. This is because both parties frequently have to exchange sensitive information in order to properly fulfill their obligations under the contract. The non-disclosure agreement should have a fixed term; in exceptional cases penalty conditions can be stipulated.

The parties should furthermore agree on the **intellectual property rights** which might result from the project. This includes patents, source codes, licenses, know-how, etc.

Finally the contract should include very general stipulations on the termination of the contract, arbitration, jurisdiction, etc.

The parties may agree on a letter of intent in the form of a Memorandum of Understanding, option contract, etc. This may be necessary if there is a considerable period between the first contacts between the parties and the conclusion of a contract. During this period there may already be negotiations including an exchange of essential information and know-how, which also might give reason for expenditures.

Particular Case: Contract Awarded by the German Aerospace Center

A particular situation arises if the **German Space Agency** within the German Aerospace Center awards a contract to industry, research organizations or universities to produce a space object. For such contracts general terms and conditions apply. The parties also

have to consider the specifications in the Federal Budget Act.

According to §55 of the Federal Budget Act, contracts assigned by the German Space Agency have to be tendered. The Agency has to consider the budget law as well as the competition regulations in order to define the right way to award the contract. Accordingly, the Agency may opt for a public tender, a restricted tender or a single tender action (in case of delivery of a product or service below the threshold level of €130 000, or €200 000 in the case of a service involving research and development) or for an open, closed or negotiated procedure (above the named thresholds). The Agency prepares a detailed specification of the services to be rendered, irrespective of which of the above-mentioned procedures is followed. This is necessary so that the applicant/tenderer can prepare an offer. Anyone interested in applying for tender studies the newspapers, official journals, professional journals as well as the supplement to the official journal of the EU because the Agency is obligated to publish its tenders in these organs. Sometimes, the Agency may simply contact certain entities and ask them to prepare an offer of services meeting the specification. But this may only be done if a tender was launched beforehand in order to prepare the list of participants from which the Agency then chooses the entities to be contacted. The tenderer has to meet the deadline for submission of the offer, as the tenderer will otherwise be excluded from further consideration. The offers received in time are checked on formal issues, after which a tender commission selects the most economical offer. Whether or not an offer is the most economical is decided using criteria established beforehand. A weighting coefficient is determined for each criterion. It is not the lowest price alone which determines the outcome.

The tenderer who bids the most economical offer is accepted to negotiate a contract on the basis of this offer. The contract is usually a **research and development contract**. This is the case if at least 25% of the contract is on research and development issues.

The contract includes a detailed description of the project, the period in which the contract is to be executed as well as an agreement about payments. The kind of payment (based on costs or expenditures) depends on the type of contractor: if the contractor uses a cost calculation, the project is remunerated on a cost basis. This applies to industry, but also to the Fraunhofer Gesellschaft and the Helmholtz Association. If the contractor does not use a cost calculation, the project is remunerated on an expenditure basis (all expenditures have to be proved by receipts). This applies to the Max Planck Society, universities and most research organizations. For remuneration on a cost basis, the (offered) market price is to be agreed. Only in the case of a single tender action or a negotiation procedure applied to the tender can remuneration (on a cost basis) be by primary costs on a fixed price basis, or by primary costs on a refund basis. The contractor bears the risk of additional cost if market price or primary costs on a refund basis was agreed. In the case of primary costs on a fixed price basis, the contractor only bears the risk of additional necessary efforts. For remuneration on an expenditure basis, the parties have to agree on a fixed price instead of a market price, or a price based on primary costs on a fixed price basis, or on an expenditure refund instead of a price based on primary cost on a refund basis.

The contractor must comply with ECSS standards in general. While these standards apply to industries only by voluntary commitment or contractual reference, the Agency is bound to these standards for all projects which are undertaken in the framework of the national space program. Hence, the Agency has to pass on these obligations to all its contractors. Finally, the contract includes stipulations on the termination of the contract, withdrawal and other general matters.

The Federal Ministry of Education and Research has specified general terms and conditions for research and technology contracts of the type of primary costs on a refund basis (BEBF 98). These terms and conditions are also used by the Federal Ministry of Economics and Technology for contracts concluded in the context of the German space program. These general terms and conditions apply by way of analogy also to contracts of the type of costs on a refund basis; for contracts based on (primary) fixed costs they apply to a limited extent. BEBF 98 deals with issues which are generally applicable to research and development contracts, in addition to the text of a particular contract. They are included in the contract by reference. Basically, they are not subject to negotiation. BEBF 98 regulates some general matters concerning how

the contract is to be executed. The chargeable costs are listed, as are the terms of invoicing and the payment itself. The terms allow for price checks by the principal. In case the contractor wants to subcontract obligations of the contract to another entity, BEBF 98 provides some rules on how the contractor can do so. The contractor has obligations to report to the principal. Furthermore, it is stipulated that all results of the research and development performance of the contractor (i.e., findings, know-how, inventions, products) be owned by the principal. But the contractor may establish an exploitation plan including a stipulation that the contractor may be entitled to exploit the result in order to receive royalties. If the employees of the contractor invent something in the course of exercising the contract, the contractor as employer has to claim the invention and to file a patent (before publishing the result/invention). BEBF 98 obliges the contractor to publish the results, but not before having filed a patent or protected any other (intellectual) property right. Finally, the principal accepts the result and produces a final report in coordination with the contractor. This includes a list of potential deficiencies. The contractor guarantees (for 12 months) that it has used state-of-the-art technology, checked the quality of materials, exercised competent and excellent performance as well as adhered to all warranted characteristics of the contract.

Particular Case: Cooperation Agreement with the German Aerospace Center

While the German Space Agency within the German Aerospace Center can only enter into a research and development contract, as described above, if there is a clear governmental need for the results, there might be cases where there is a need of the result for both Germany and the contractor. Here the parties choose a **public–private partnership** agreement [9.4.51]. Such agreements are explicitly mentioned in the German space program as a specific and desired means to achieve the aims of the program. Public–private partnership agreements were used by the Agency in the past mainly to execute the remote sensing program.

A public–private partnership contract reflects the special interest of both parties in the result. As in any other contract, the obligations and responsibilities of both parties have to be described as extensively as possible. Again, the more technical details are annexed to the contract. This includes definition of the risks which each of the parties bear in case of non-achievement of the results. While the ownership of the results (especially patents) is with the principal in a research and development contract, the regulations in a public–private partnership are different. Here there is no clear distinction made between principal and contractor; that is, which party has to develop a result against payments by the other party. This is why the parties explicitly have to agree on who should be the owner of which result, on how to proceed with filing any results if this is possible, and who has an additional right to use (exclusively or non-exclusively) these results. A public–private partnership agreement also has to include stipulations on how payments between the parties are made. As there might be damage to the parties while the contract is executed, they should agree on liability clauses. It is common understanding in many research and development contracts on outer space projects that the parties agree on a cross-waiver of liability. This means that none of the parties is allowed to claim damages from the other party. In Germany, such a cross-waiver has to be limited to damages which are caused by negligence or carelessness. For damages caused by willful misconduct or gross negligence, the other party does not have to refrain from claiming damages. The cross-waiver is intended to limit the (financial) risks of both parties. Finally, the public–private partnership agreement contains general stipulations on nondisclosure, publication, termination and arbitration or litigation.

Launch of a Space Object

A satellite or probe cannot be operated in outer space before it is launched. This transport service can be purchased in the marketplace.

The parties to a **launch service agreement** are usually the launch service provider and the operator of the object to be launched. But the operator can also negotiate the above-mentioned contract for manufacturing the satellite in such a way that the manufacturer's obligation is not only to build and provide the satellite, but also to launch and place the satellite in orbit (a **delivery in orbit** or **turnkey contract**). Then, it is not until the positioning took place that the manufacturer has met all its obligations

toward the principal (operator). It is the manufacturer who has to enter into a launch service agreement with the launch service provider.

The main content of a launch service agreement [9.4.52] is to **attribute the risk of a launch failure** or wrong positioning to one of the parties of the agreement. This is done by several stipulations within the agreement, which have to be read together.

The main contractual obligations are the following: While the operator is obliged to pay a certain amount to the launch service provider (depending on the rocket used, currently roughly $10 000–20 000 per kilogram), the launch service provider is obliged to launch the object. The agreement has to specifically define what a "launch" is: that is, when the launch service provider has fulfilled its obligations under the agreement. This is important, as at this point it is not the launch service provider but the operator who bears any risk of failure or wrong positioning. The parties are free to stipulate that the launch took place when the rocket was ignited and took off. But they may also stipulate that the launch took place when the designated orbit was reached, the commissioning phase was finalized, or the satellite started operation.

The risk of failure is with the launch service provider to achieve what has been agreed. But the launch service agreement differs from many other contracts insofar as this risk is limited in two ways. First, a launch service agreement typically includes a **best effort** clause. This specifies that the launch service provider is not obliged to guarantee the result (i.e., positioning, or whatever has been agreed) but only obliged to make a best effort to achieve this result. Only if a standard type of satellite bus is launched repeatedly with the same rocket type might there be a reason to agree with the launch service provider that the result is guaranteed. If the agreement does not include the one or the other clause, the launch service provider is usually only obliged to provide a best effort, based on the applicable laws. Second, the agreement typically includes a restriction, according to which the parties agree on a **cross-waiver of liability**. The launch service provider only bears the risk of damages which occur on its side (especially damage to the rocket), while the operator bears the risk of the damages occurred on its side (especially damage to the object which was going to be launched). Such a cross-waiver is usually independent of any stipulation made on the basic risk attribution. But a cross-waiver might be void if one of the parties caused the damage through gross negligence or willful misconduct.

As a result, the launch service provider typically bears the risk that it does not provide the best effort to achieve the agreed result; in case damage occurs, although the launch service provider has provided the best effort, it bears the risk for damage occurring to itself as well as to any third party. But the launch service provider does not bear the risk for damage caused to the operator, especially not damage to the satellite which was going to be launched.

Compared to any other type of contract on transport, the risks shared between the parties are shifted toward the operator of the object to be launched. This is because of the likelihood of damages which might occur in the launch phase, but also to the uniqueness of each agreement, since almost every rocket has to be adapted for the specific satellite bus type used. But the operator has other means within the agreement to shift some risk back to the launch service provider. The latter can agree with the operator for an unpaid re-launch. Hence, the operator only bears the costs for rebuilding the satellite, but does not need to buy a launch rocket for the second attempt. The operator also should try to obligate the launch service provider to accept reviews and procedures for accepting and approving intermediate services performed by the launch service provider as well as to admit employees of the operator to the site of the launch service provider while the launch is being prepared.

The remaining risks for both parties may be (partly) shifted to an insurer.

In addition to these stipulations on risk sharing, a launch service agreement contains several other elements: the parties agree on definitions of essential terms used in the agreement, in order to achieve a common understanding (e.g., failure, partial loss, total loss); the launch schedule as well as the launch window will be agreed upon, as well as milestones for payment (e.g., based on the progress of launch preparations) and rules for postponement of the launch. The parties should also agree on which party is responsible for applying for authorizations (e.g., import and export, launch license, frequency allocation, registration of the satellite as well as any rocket stage which might remain

in orbit for some time). It is suggested that a stipulation on how to inform all launching states is also included. A launch service agreement often includes the clarification that ownership of the object which is going to be launched is not affected by its integration in the rocket. Questions of intellectual property rights as well as the transfer of know-how might also be of importance to the parties. The parties might agree on non-disclosure of information and termination of the contract (e.g., the right of the operator to terminate the contract if the rocket type which is intended to be used for the launch fails in another launch which takes place right before the launch which is the subject of the agreement).

Operation of a Space Object

Depending on the specific space mission, the operation of a satellite, probe, etc., gives rise to many legal issues. Three of them are mentioned here as examples.

If the operator is not able or does not want to transmit the signals for **telemetry, tracking and control**, the operator can contract this service to the operator of a ground station. Even if this is not intended for the whole mission, it is usually done for positioning a satellite. This is because the coverage of the central ground station might not be sufficient for contacting the satellite as often as necessary to conduct all foreseen maneuvers and tests. Also, a ground segment can be contracted in order to download satellite data because the cache capacity available on-board the satellite is restricted. This is the case for many remote sensing activities. A contract with a ground segment operator must specify the obligations of both parties (transmission of signals or reception of data), the handling of signals (authentication) and data (encryption). The parties should also agree on backups in case the ground segments contracted are not able to send signals to the satellite. Finally, the parties should agree on who is responsible for applying for the allocation of necessary frequencies.

The operator might need to conclude **sales contracts for products and services** (e.g., remote sensing data, navigation signals, telecommunication services). Again, the product or service to be delivered should be described as precisely as possible, because the clients may raise liability claims for breach of warranty. For some services the service provider may be able to negotiate a disclaimer of warranty; for example, for the case that there is an interruption in supplying navigation signals which are transmitted for no or little remuneration. Especially in the domain of navigation services, the parties need to focus on consequential damages caused by the malfunction of the signals. If the service provided also includes the development of products, such as remote sensing data products, the parties have to conclude user (license) agreements for them.

If a satellite operator is not in need of all the space on a satellite bus, the operator may agree with other service providers to integrate third-party payloads on the satellite bus. This might be the case if a telecommunications operator only requires some of the transponders on a satellite for the operator's own use and makes the others available to another provider. This may constitute a funding source for the satellite operator (see below), which means that the operator will enter into a **rental or leasing contract**. The contract should especially stipulate the conditions under which the lessee or tenant may control the leased transponders irrespective of and independent from the other transponders or under which conditions the lessee or tenant may change telemetry, tracking and command signals in order to adapt the operation of the whole space segment to the needs of the lessee or tenant's own services. The transponders leased to another party need to be described and located as precisely as possible in order to differentiate them from the transponders owned by the operator or any other person. If a transponder fails in the course of the mission, it has to be clear who is the owner or lessee of that specific transponder, and the lessor and the lessee have to accordingly agree on who bears the risk of failure and to which extent this risk is borne.

Insurance for Space Activities

The insurance market is now able to provide a number of different kinds of **insurance** for space activities.

Any person involved in a space mission has to consider whether they want to bear the technical and commercial risks remaining with them according to the contracts concluded, or whether they want to cover these risks by taking out an insurance policy or by using their savings (which would be **self-insurance**). There

might be an area where it is compulsory to take out an insurance policy: some national legislation requires persons to have third-party liability insurance.

In Germany, there is for the time being no such **compulsory third-party liability insurance**. But any German operator has to realize that also foreign legislation requiring such a compulsory insurance might have an effect. If for example a German satellite is launched from the USA, Russia or Australia, the launch (including the payload satellite) has to be compulsorily insured against third-party liability claims. For Ariane launches, compulsory third-party insurance is required quasi by law. If a German satellite is operated by a US, Russian or UK national, the operation also has to be insured against third-party liability claims. In the USA, such insurance has to have a maximum coverage equivalent to about €415 million (or the maximum probable loss), in Australia the equivalent of roughly €450 million (but at least the maximum probable loss), in Russia the equivalent of between €80 and €250 million, and in the United Kingdom the equivalent of roughly €145 million. Ariane launches have to be insured against third-party liability claims to a maximum of €60 million. Similar requirements exist in Hong Kong, Ukraine, South Africa, Brazil and Belgium.

Other insurance may be taken out on a voluntary basis. Today, almost all aspects of a space mission are insurable: damage caused to someone's own property, damage to contractors and damage to third parties. There is insurance coverage for damage occurring in the course of manufacturing a satellite, its transport to the launch facility, its integration into the rocket, its launch, its positioning as well as its operation in outer space. To some extent, also the retention of an inoperable satellite in outer space (as debris) as well as the reentry of a satellite into the Earth's atmosphere might be insured. The insurance might cover technical failures, delays in operation, or loss of income.

The parties have to define in the insurance policy the insured event. This includes agreement on when the parties assume damage to a satellite to be an interruption of business, a partial loss, or a total loss (which can already be the case if 51% of the service is unavailable). Some risks (e.g., *force majeure*) might be excluded by the parties. The parties have to determine the amount of coverage, that is the sum insured and paid by the insurer in case the insured event takes place. Based on the amount of coverage, the parties agree upon the insurance rates which are to be paid. As the insurance market at the moment is very unstable and volatile, specific figures are very much dependent on the moment when the insurance contract is concluded as well as on the peculiarities of the contracted space mission. Finally, the parties define the duration of coverage. For the operation of a satellite the usual duration is currently 12 months, subject to renewal.

Any person intending to conclude insurance must thoroughly assess the **risks** remaining with that person. If, for example, the operator of a satellite has ordered a satellite from a manufacturer, that manufacturer is responsible for delivering the satellite as specified in the contract. Damage or malfunctions in this area are the risk of the manufacturer. If, for example, the operator has concluded a launch service agreement, the launch service provider has to take the risk of a launch failure. But this risk is, as seen above, limited in certain ways. If the launch service agreement stipulates that the launch has taken place when the rocket has lifted off, and the failure happens while the satellite is being positioned, the operator should have insured this. If the launch service agreement contains a cross-waiver of liability, the operator should insure for damage to the satellite also occurring in the launch phase.

The negotiation of an adequate and balanced insurance policy is very much dependent on the particularities of every single case [9.4.53], [9.4.54].

Many (technical) details of the satellite, its payloads and the rocket have to be transmitted to the insurer for calculating the amount to be covered as well as the insurance rate. If the insurer is based in a foreign state, the transmittal of such information might be technology listed as a dual-use item, which can be subject to export control regulation.

Financing Space Activities

Before the first income can be generated (e.g., by providing telecommunication services or by selling remote sensing satellite data), the operator must invest a lot of resources. The satellite usually will be built by a manufacturer, launched by a launch service provider and may be tracked and commanded by a ground segment operator. They all have to be paid.

Often, at the beginning of such a space mission, the provider does not have many funds at its disposal. Therefore, the provider will enter into contracts with people who can provide credit capital. This can be done with a **conditional sales contract** concluded with the manufacturer: the payment of the manufacturer will be settled by installments, the satellite will be delivered to the operator upon completion, but ownership of the satellite remains with the manufacturer until the last installment is paid by the operator. A credit agreement can also be in the form of a **leasing contract** with a bank. Instead of the operator, the bank purchases the satellite from the manufacturer. Hence, the bank is owner of the satellite. The bank will provide the right of use to the operator, and the operator has to pay leasing installments in return. Finally, a general **credit agreement** can be concluded. A credit agreement obliges the bank to provide a certain amount of payment to the operator, while the operator transfers several security rights to the bank (project financing). This includes an agreement according to which the bank becomes the owner of some assets (e.g., the satellite) and transfers to the operator the right to use these assets (transfer by way of security). In addition the operator may pledge stakes in its corporation or entitlements of contracts to the bank, such as the right to use TT&C, the right to replace the operator in the launch service agreement, insurance rights or license agreements [9.4.55].

If the operator cannot repay the credit to the bank, the bank can make use of the security interest granted to it: if the bank has agreed on a general credit agreement the bank can liquidate the rights resulting out of the contracts or stakes pledged to it. If ownership in the satellite was transferred to the bank, the bank may terminate the right of the operator to use the satellite. It may then sell the satellite to a third party. If the manufacturer has agreed upon a conditional sale or leasing agreement, it can terminate this agreement. It may then transfer the right to use the satellite to a third party or even sell the satellite.

Both the bank and the manufacturer, who are owners of a satellite by way of the transfer of security interests, have to consider that the **exercise of remedies** (e.g., to sell the satellite) can proceed only according to law. The parties cannot deviate by contract. The bank faces difficulties especially with regard to mobile equipment, such as satellites: the remedies may only be exercised by the laws of the state, where the equipment is located at the moment of default (i.e., the operator is no longer able to pay the credit installments). The bank and the manufacturer will have to take this uncertainty into account when defining the amount of credit paid in respect to the amount of the installments. Therefore the parties should know into which states the equipment might be brought. For satellites, the places of manufacturer, transport and launch are important.

From a legal point of view, there will be difficulties in defining the law applicable to the exercise of remedies when the satellite has been launched and placed in outer space. According to the above, the laws of the state where the equipment is located applies. But as also noted above, outer space is international territory (Art. II of the Outer Space Treaty), which also excludes the application of national laws. Some scholars argue that for the exercise of remedies the laws of the state apply in which the satellite has been registered. This is because this state of registry retains jurisdiction over and control of the object according to Art. VIII of the Outer Space Treaty. This "jurisdiction and control" includes application of the national laws, so they argue. According to other scholars, the laws of the state where the main ground segment (control station) is located apply [9.4.56].

This legal uncertainty may be overcome in the medium term. The **UNIDROIT** Convention on security interests in high-value mobile equipment (Cape Town Convention) provides for a harmonization of the national laws on security interests. States ratifying this Convention are called to amend their national laws according to the rules of the Convention. The more states have done so, the less uncertainty remains for banks concerning how to exercise remedies in foreign states. The Cape Town Convention today only applies to aircraft, engines and helicopters. It will soon be applicable to railways and railway wagons. And it might become applicable to satellites and other space objects through ratification of a protocol on space assets [9.4.57]. This protocol is still under negotiation and some important questions (e.g., to which assets the protocol applies) are still unanswered.

9.4.2.4 Liability Risks

Every participant in a space mission should be aware of liability risks in the course of the mission. These should be considered at the very beginning. In addition, it has to be decided which of these risks should be insured.

A space mission is subject to ordinary liability risks, just as any other activity. But there are additional liability risks which have to be taken into account.

First, the activity is subject to the typical liabilities resulting from the laws as well as from all concluded contracts. If a contracting party does not meet its **contractual obligations** the contracting partner may claim compensation for damage. The same applies if a contracting partner does meet its obligations, but incompletely or incorrectly or only after a delay. If a **third party is damaged** in the course of the mission, this third party may claim for compensation of damages according to the law (tort law). The party damaged is entitled to compensation for every financial loss which occurred, irrespective of whether the claim is based on a contract or on the laws. The party who caused the damage has to reestablish the situation as it was before the damaging event took place. This compensation may be done by repairing the damage. But usually compensation takes place **by payment** of a certain sum of money.

Some contractual liabilities may be excluded by the parties. A cross-waiver of liabilities for damages which were not caused through gross negligence or willful misconduct is very typical in the context of research and development contracts.

Basically, a contractual liability can only be claimed if the damage is based on the fault of one contracting party. Fault is given if the damage was caused willfully or by negligence. The legislation may, however, specify that for certain activities liability is also given without fault. This is called **strict (absolute) liability**. The reason for such a legislative decision may be that an activity is deemed to be basically dangerous. A strict liability implies a higher liability risk: a person damaged does not need to prove that the person responsible for the damage has caused the damage by fault. A strict liability exists, for example in air traffic law for the operation of an aircraft (§3 of the German Air Traffic Act), but not for space activities. However, some other states have implemented in their national legislation a strict liability for the launching and reentry of space objects. The reasoning is that a similar concept applies in international law, since the launching state is absolutely liable for damage caused by space objects on the Earth or to aircraft in flight, see Art. II of the Liability Convention.

Many space activities are conducted under international cooperation. This may raise the question of which national law is applicable to liability issues. Every contract should therefore include a clause agreeing on the applicable law. If the parties have not agreed on this, the law of the place where the contractual obligation was accomplished is applicable. If a third party is damaged in the course of a space mission and the context is international, the parties cannot agree on the applicable law. The claims are handled according to the laws of the state in which the damage occurred.

For damages caused to third parties some specifics have to be taken into account. A person damaged may directly claim for compensation from the person who caused the damage. In addition, a liability of the launching state exists according to the international treaties (Art. VII of the Outer Space Treaty, Arts. II–IV of the Liability Convention). The person damaged may choose whether to claim against the person who caused the damage or against a launching state. The liability of the launching state is not linked to the factor which caused the damage, but only to the participation of that state in the event of the launch. This international liability also exists when the damage is caused years later in the course of operation or reentry of the object into the Earth's atmosphere. And this international liability also exists if the damage was caused by the operator. If the person damaged chooses to claim against a launching state, this state might be interested in taking recourse against the operator or any other person who caused the damage. This can be done only if there is a legal basis in the national laws. This is the case in some states (e.g., Russian Federation, USA, United Kingdom and Australia). Most of these states have at the same time limited this recourse to a certain amount of money, typically the amount of the compulsory third-party liability insurance of the operator. This leads to insurability of the activity and also provides some financial protection for the

national industry. In Germany there is actually no legal basis for such recourse.

For any person launching or operating a space object, the additional international liability of the launching states has a positive effect. It can be of advantage for the operator who caused damage to a third party if this third party does not claim compensation against the operator but against a launching state. This launching state may indeed take recourse against the operator. But this **recourse** is often limited. Then the operator and the launching state(s) share the liability. Otherwise, in case there is a direct claim of the third party against the operator, the operator has to fully compensate the third party.

Another specific aspect of liability issues in outer space is the treatment of **space debris**. Space debris consists of human-made objects which are in outer space but which have lost their function of using or exploring outer space. Debris may also include former parts of a space object. The avoidance and mitigation of space debris is an important and current topic in many fora. Many approaches are taken. The most famous might be the approach of the Inter-Agency Space Debris Coordination Committee. It has developed rules which are widely used as technical standards.

If damage occurs despite all efforts to avoid or mitigate debris, the question arises whether or not the (former) operator is liable for such damage. Liability does not end with the termination of operation. Space debris is also regarded as an object which was once launched for the purpose of using or exploring outer space. The launching states are liable regardless of the functional status of the object. Their liability is given because of their involvement at the very moment of the launch. If space debris later causes damage and if this damage even was caused after reentry into the Earth's atmosphere or to aircraft in flight, the launching state is absolutely liable (see Art. II of the Liability Convention). Otherwise it is fault liability (Arts. III, IV of the Liability Convention), but fault is already given if technical standards for avoidance or mitigation of debris were not followed, either willfully or by negligence.

To the same extent the (former) operator is also liable for damages caused by space debris. If the (former) operator has willfully or by negligence omitted to consider measures to avoid damage (e.g., to put the object into a graveyard orbit or to bring the object to a controlled re-entry), this may cause liability.

Any operator who has the right to use frequency and orbital position because of a transfer of such rights by the Federal Network Agency are subject to a condition that sufficient fuel has to be retained in order to allow orbital maneuvers to terminate operation. Satellites in a geostationary orbit should be brought into a graveyard orbit or caused to reenter the Earth's atmosphere according to 3.7 of Administrative Order 8/2005.

9.4.3 Space Activities in the Framework of the EU and ESA

European activities in outer space have never been exclusively national, bilateral or trilateral activities. There has always been a European dimension. This European dimension is also of relevance for every private entrepreneur, university or research institution. Every decision taken in the framework of the ESA as well as in the framework of the EU has implications for their activities. Especially those decisions which are of a political and programmatic character and which define their mid-term perspectives can influence the strategy of such non-governmental entities. It might be interesting for them to orientate according to these governmental perspectives, as there might be financial support forthcoming in the form of grants. Since these decisions define areas in which the EU or ESA might assign contracts to national industry, university or research organizations in the future, it is recommended that these decision-making processes on political and programmatic issues at the European level be closely followed to the same extent that the same processes at the national level are followed.

Both the EU and ESA have meanwhile become actors in outer space. But both have to be strictly differentiated, also because of their different histories. They will therefore be described separately.

The main decisions within ESA are made by the Council at ministerial level. This Council is convened every two to four years. It is within this Council that the ESA member states agree on the basic political principles of the organization's space policy. At the

same time the member states designate the amount of their participation within the (optional) program lines of ESA; that is, how much money states will transfer to ESA for executing the programs. As ESA is based on the principle of geo-return, by deciding on the amount of money to be transferred to ESA, the member states decide at the same time about the percentage of money which is going to be retransferred to their national industry (universities and research organizations). This principle basically says that if a member state pays a certain percentage of money for a defined program, the same percentage value of contracts has to be concluded with entities of that same state for executing this program.

The main decisions within the EU are made by the Council and the Parliament or within the European Commission. Outer space activities will be explicitly mentioned for the first time in the basic treaties of the EU when the EC Treaty as amended by the **Lisbon Treaty** enters into force. Therein the competence for carrying out space activities (especially the definition and implementation of programs) will be shared among the EU and the member states. The EU is called on to draw up a European space policy and to establish the necessary measures, which may take the form of a European space program, excluding any harmonization of the laws and regulations of the member states.

On another level, the European Commission (which is, broadly speaking, the Executive) has decided on some basic principles on space published in its **White Paper** entitled "Space: a new European frontier for an expanding Union. An action plan for implementing the European space policy." The White Paper implements European space policy (which is a policy at the EU level and should not be confused with the European Space Policy which recently was agreed between the EU and ESA, see below). Global Monitoring for Environment and Security (GMES) as well as Galileo have a central role within the White Paper, accompanied by issues on digital divide, outer space for common security and defense policies, monitoring conflict areas from outer space and monitoring compliance with international treaties and agreement (e.g., treaties on disarmament). Cooperation between the EU and ESA as well as with the Russian Federation is explicitly mentioned within the aims of the policy. Besides naming these aims, the White Paper also defines key elements for achieving them: Europe has to achieve strategic independence from other spacefaring nations, especially with regard to parts and compliances for space objects as well as with regard to access to space; the focus is to be on technologies which are of special importance for future developments, and the exploration of outer space and scientific education are considered to be essential elements. Also on the level of the European Commission, the **7th Framework Program** should be mentioned here. This program started in 2007 and defines the areas in which the European Commission commissions (outer space) research activities to European industries, universities and research organizations. Tendering for these contracts is in full and open competition.

Cooperation between the EU and ESA has grown in recent years, despite their different histories and tasks, almost conflicting industrial policies and some differences in their member states [9.4.58]. This was decided by the Council of ESA as well as by the Council of the EU. Based on these decisions, cooperation was formalized in common resolutions of the ESA and EU Councils on a common European Space Policy and on a **Framework Agreement** for further cooperation. The core stipulation of the Framework Agreement is basically that both continue to execute their tasks based on their own internal procedures, but undertake cooperation in scientific activities, technologies development, remote sensing, navigation, telecommunication, human space activities, microgravity research and launching activities. In practical terms, the EU participates in the optional ESA programs of Galileo. Galileo is the first huge program in which both bodies cooperate. The second program will be GMES, which is currently being executed by ESA only, but is intended to be transferred to the EU in late 2008. Finally, the Framework Agreement is the basis for the development of a common European Space Policy. In preparation of this policy, a new body was created, the **Space Council**, which is broadly speaking a union of the Councils of the EU and ESA.

Whether and to what extent the cooperation between the EU and ESA will be extended to include organizational issues is still written in the stars. Different models are being discussed. These discussions are going to be very specific, but in broad terms whether

ESA might for example become an agency of the EU, or, at the other extreme, whether the EU becomes a member "state" of ESA, is up for discussion [9.4.60].

Bibliography

[9.4.1] Von Kries, W., Schmidt-Tedd, B., Schrogl, K.-U. *Gründzüge des Raumfahrtrechts*. Munich: Verlag C.H. Beck 2002, 31f.

[9.4.2] Bueckling, A. *Der Weltraumvertrag*. Cologne: Carl Heymanns Verlag, 1980.

[9.4.3] Creola, P. *Raumfahrt und Völkerrecht*. Zurich: Polygraphischer Verlag, 1970.

[9.4.4] Hobe S., Schmidt-Tedd B., Schrogl, K.-U. (eds.) *Current Issues in the Registration of Space Objects*, Proceedings of Project 2001 Plus Workshop. Cologne: Deutsches Zentrum für Luft- und Raumfahrt, 2005.

[9.4.5] Lafferranderie, G. Jurisdiction and Control of Space Objects and the Case of an International Intergovernmental Organization (ESA). *Z. Luft- Weltraumrecht*, 228–242, 2005.

[9.4.6] Schmidt-Tedd, B., Gerhard, M. Registration of Space Objects: Which Are the Advantages for States Resulting from Registration? In Benkö, M., Schrogl, K.-U. (eds.), *Space Law: Current Problems and Perspectives for Future Regulation*. Utrecht: Eleven International Publishing, 2005, 121–140.

[9.4.7] United Nations. Practice of States and International Organizations in Registering Space Objects. *UN Doc. A/AC.105/C.2/L.255*.

[9.4.8] Rodrigues, N. The United Nations Register of Objects Launched into Outer Space. In Hobe S., Schmidt-Tedd, B., Schrogl, K.-U. (eds.), *Current Issues in the Registration of Space Objects*, Proceedings of Project 2001 Plus Workshop. Cologne: Deutsches Zentrum für Luft- und Raumfahrt, 2005, 25–35.

[9.4.9] United Nations. Online Index of Objects Launched into Outer Space. http://www.unoosa.org/oosa/en/SORegister/index.html.

[9.4.10] Bittlinger, H. *Hoheitsgewalt und Kontrolle im Weltraum*. Cologne: Carl Heymanns Verlag, 1988.

[9.4.11] Bueckling, A. Das Übereinkommen vom 29.3.1972 über die völkerrechtliche Haftung für Schäden durch Weltraumgegenstände. *Versicherungsrecht*, 389–396, 1977.

[9.4.12] Malanczuk, P. Haftung. In Böckstiegel, K.H., *Handbuch des Weltraumrechts*. Cologne: Carl Heymanns Verlag, 1991, 755–803.

[9.4.13] Wins, E. *Weltraumhaftung im Völkerrecht*. Berlin: Duncker und Humblot, 2001, 62–65.

[9.4.14] Gerhard, M. *Nationale Weltraumgesetzgebung*. Cologne: Carl Heymanns Verlag, 2002, 126–129.

[9.4.15] United Nations: Resolution 59/115 zur Anwendung des Konzeptes des "Startstaates," December 10, 2004.

[9.4.16] Gerhard, M. Transfer of Operation and Control with Respect to Space Objects – Problems of Responsibility and Liability of States. *Z. Luft- Weltraumrecht*, 571–581, 2002.

[9.4.17] Lee, R.J. Creating an International Regime for Property Rights under the Moon Agreement. IISL Proceedings of the 42nd Colloquium on the Law of Outer Space, 1999, pp. 409–418.

[9.4.18] Kimminich, O., Hobe S. *Einführung in das Völkerrecht*, 7. Auflage. Tübingen: Francke, 2000, 189–190.

[9.4.19] Benkö, M., Schrogl K.-U. The UN Committee on the Peaceful Uses of Outer Space: Adoption of a Declaration on "Space Benefits." *Z. Luft- Weltraumrecht*, 228–248, 1997.

[9.4.20] Benkö M., Schrogl K.-U. The UN Committee on the Peaceful Uses of Outer Space: Adoption of a Resolution on Application on the Concept of the "Launching State" and Other Recent Developments. *Z. Luft- Weltraumrecht*, 58–72, 2005.

[9.4.21] Baumann, I. *Das internationale Recht der Satellitenkommunikation*. Frankfurt a.M.: Peter Lang, 2005, 267–330.

[9.4.22] Tegge, A. *Die Internationale Telekommunikations-Union*. Baden-Baden: Nomos, 1994.

[9.4.23] Bollweg, H.-G., Gerhard M. Sicherungsrechte an Luftfahrtausrüstung und Weltraumeigentum – Die Entwürfe einer UNIDROIT/ICAO – Konvention über internationale Sicherungsrechte an beweglicher Ausrüstung und der Protokolle über Luftfahrtausrüstung und Weltraumeigentum. *Z. Luft- Weltraumrecht*, 373–396, 2001.

[9.4.24] de Fontmichel, A. Commentaires sur l'avant projet de Protocole sur les questions spécifique aux matériels d'équipement spatial en projet de Convention d'UNIDROIT relative au garanties internationales portant sur des matériels d'équipement mobiles. *Z. Luft- Weltraumrecht*, 526–552, 2001.

[9.4.25] Nagel, K.-F. Das neue Regulierungsübereinkommen über die Internationale Raumstation. *Z. Luft- Weltraumrecht*, 143–149, 1998.

[9.4.26] Bourbonniere, M. National-Security Law in Outer Space: The Interface of Exploration and Security. *J. Air Law Commer.*, 3–62, 2005.

[9.4.27] Böckstiegel, K.H., Benkö, M., Hobe S. *Space Law: Basic Legal Documents*, Instalment 9. Utrecht: Eleven International Publishing, 2004.

[9.4.28] Bundesnetzagentur: Anträge und Ausfüllhinweise. http://www.bundesnetzagentur.de/enid/4006a1859b2fdbccc12f4829dff7b3fb,0/Satellitenfunk/Antraege_und_Ausfuellhinweise_db.html, November 22, 2005.

[9.4.29] Bundesnetzagentur: Frequenznutzungsplan. http://www.bundesnetzagentur.de/media/archive/1820.pdf, May 2006.

[9.4.30] Baumann I. *Das internationale Recht der Satellitenkommunikation*. Frankfurt a.M.: Peter Lang, 2005.

[9.4.31] Baumann, I., Gerhard M. Neuregelung des Verfahrens zur Anmeldung eines Satellitensystems bei der ITU und zur Übertragung deutscher Orbit- und Frequenznutzungsrechte. *Z. Luft- Weltraumrecht*, 87–99, 2006.

[9.4.32] Creydt, M. US-Exportrecht und Satellitenindustrie. *Z. Außenwirtsch. Recht Prax.*, 453–457, 2002.

[9.4.33] Bundesamt für Wirtschaft und Außenhandel: Ausfuhrliste: http://www.ausfuhrkontrolle.info/vorschriften.php, June 2007.

[9.4.34] Bundesamt für Wirtschaft und Ausfuhrkontrolle: Formulare. http://www.ausfuhrkontrolle.info/formulare.php, June 2007.

[9.4.35] Cloppenburg, J. Jüngste Entwicklungen im U.S.-amerikanischen Außenwirtschaftsrecht – Die Regulierung von Hochtechnologieexporten und ihr Einfluss auf die betroffenen Wirtschaftszweige am Beispiel der amerikanischen Satellitenindustrie. *Z. Luft- Weltraumrecht*, 510–525, 2001.

[9.4.36] Von Kries, W., Schmidt-Tedd, B., Schrogl, K.-U. *Grundzüge des Raumfahrtrechts.* Munich: Verlag C.H. Beck, 2002, 93–101.

[9.4.37] Gerhard, M. *Nationale Weltraumgesetzgebung.* Cologne: Carl Heymanns Verlag, 2002.

[9.4.38] Gerhard, M., Schmidt-Tedd, B. Regulatory Framework for the Distribution of Remote Sensing Satellite Data: Germany's Draft Legislation on Safeguarding Security Interests. IISL Proceedings of the 48th Colloquium on the Law of Outer Space, 2005, pp. 45–54.

[9.4.39] http://www.bundesrecht.juris.de/bundesrecht/satdsiv/gesamt.pdf, June 2007.

[9.4.40] Department of Commerce: Licensing of Private Remote Sensing Systems, 15 CFR 960. http://www.licensing.noaa.gov/CRSLRegs04-25-06.pdf.

[9.4.41] Parliament of Canada: An Act Governing the Operation of Remote Sensing Space Systems. http://www.parl.gc.ca/PDF/38/1/parlbus/chambus/house/bills/government/C-25_4.PDF.

[9.4.42] Stojak M. L. Regulatory Framework for Commercial Remote Sensing Satellite Systems. IISL Proceedings of the 47th Colloquium on the Law of Outer Space, 2004, *IAC-04-IISL.1.02.*

[9.4.43] Bundesministerium für Bildung und Forschung: Deutsches Raumfahrtprogramm. http://www.dlr.de/rd/dokumente/drp.pdf, May 2001.

[9.4.44] Heymer, W. Einige Gedanken zur Entwicklung des Begriffs Weltraumgegenstand. *Z. Luft- Weltraumrecht*, 71–80, 1973.

[9.4.45] Hintz, M. Weltraumgegenstände. In Böckstiegel, K.H. *Handbuch des Weltraumrechts.* Cologne: Carl Heymanns Verlag, 1991, 157–203.

[9.4.46] Swedish National Space Board: Register of Swedish Objects Launched into Outer Space. http://www.snsb.se/register_of_swedish_objects_english.shtml.

[9.4.47] British National Space Centre: UK Registry of Space Objects. http://www.bnsc.gov.uk/assets/channels/about/ukregistry.pdf.

[9.4.48] Australian Government: Register of Australian Space Objects. http://www.industry.gov.au/space.

[9.4.49] Space and Advanced Technology: US Space Objects Registry. http://www.usspaceobjectsregistry.state.gov.

[9.4.50] Schmeißer, N., Zirkel, M.P. Forschungs- und Entwicklungsverträge – Rechtliche Einordnung und vertragliche Gestaltung. *Monatsz. Dtsch. Recht*, 849–853, 2003.

[9.4.51] Späth, L., Michels, G., Schily, K. *Private Public Partnership – Die Privatwirtschaft als Sponsor öffentlicher Interessen.* Munich: Droemer, 1998.

[9.4.52] Knittelmayer, N. *Der kommerzielle Startdienstleistungsvertrag.* Baden-Baden: Nomos Verlagsgesellschaft, 1998.

[9.4.53] Gaubert, C., Moysan, S. L'assurance spatial. *Rev. Fr. Droit Aér. Spat.*, 249–265, 2003.

[9.4.54] Moysan, S. The Insurance Point of View. In Hobe, S., Schmidt-Tedd, B., Schrogl, K.-U. *et al.* (eds.), *Towards a Harmonised Approach for National Space Legislation in Europe.* Cologne: Deutsches Zentrum für Luft- und Raumfahrt, 2004, 113–117.

[9.4.55] Höpfl M. *Rechtsrahmen für die wirtschaftliche Nutzung des Weltraumes durch Privatunternehmen.* Würzburg: Ergon Verlag, 2003.

[9.4.56] Creydt, M. *Die Besicherung von Weltraumvermögenswerten – Ein neues einheitliches internationales Sicherungsrecht und dessen Vergleich zum US-amerikanischen Mobiliarsicherungsrecht.* Frankfurt a.M.: Peter Lang, 2007.

[9.4.57] Larsen P.B. Future Protocol on Security Interests in Space Assets. *J. Air Law Commer.*, 1071–1105, 2002.

[9.4.58] Hobe S., Kunzmann, K., Reuter, T. *Forschungsbereicht ESA–EU, Rechtliche Rahmenbedingungen einer zukünftigen kohärenten Struktur der europäischen Raumfahrt.* Berlin: Lit Verlag, 2006.

[9.4.59] http://ec.europa.eu/enterprise/space/doc_pdf/esp_comm7_0212_en.pdf

[9.4.60] von der Dunk, F.G. Towards One Captain on the European Spaceship – Why the EU Should Join ESA. *Space Policy*, 83–86, 2003.



Acronyms and Abbreviations

AAAF	Association Aéronautique et Astronautique de France
AAM	Auto-Acquisition Mode
AC	Alternating Current
ACRV	Assured Crew Return Vehicle
ACS	Advanced Camera for Surveys; Atmosphere Control and Supply; Attitude Control Subsystem
ACT	Attitude Control Thruster; ATV Crew Trainer
ACU	Antenna Control Unit
ACWP	Actual Cost of Work Performed
AD	Analog–Digital
ADN	Ammonium Dinitramide
ADP	Acceptance Data Package
AE	Approach Ellipsoid
AECMA	Association Européenne des Constructeurs de Matériel Aérospatial (European Association of Aerospace Industries)
AFSK	Audio Frequency Shift Keying
AGC	Automatic Gain Control
AI	Approach Initiation
AIAA	American Institute of Aeronautics and Astronautics
AIAE	Asociación de Ingenieros Aeonáuticos de España (Center in Madrid, Spain)
AIDAA	Associazione Italiana di Aeronautica e Astronautica (Center in Rome, Italy)
AIT	Assembly, Integration and Testing
AIV	Assembly, Integration and Validation
AKM	Apogee Kick Motor
ALU	Arithmetic and Logic Unit
AM	Amplitude Modulation
A-MSS	Aeronautical Mobile Satellite Service
AO	Announcement of Opportunity
AOCS	Attitude and Orbit Control System
AOS	Acquisition Of Signal
APDS	Androgenous Peripheral Docking System
APM	Attached Pressurized Module
APS	Active Pixel Sensor
APSK	Asymmetric Phase Shift Keying
APTC	Ambient Pressure Thermal Cycling
AR	Acceptance Review
ARD	Atmospheric Reentry Demonstrator
ARES	Air Revitalization System
ARTA	Ariane Research and Technology Accompaniment
ASI	Agenzia Spaziale Italiana (Italian Space Agency)
ASS	Amateur-Satellite Service
AST	American Segment Trainer
ASTM	American Society for Testing and Materials
ASTRA	family of geostationary satellites (SES Global SA)
ATB	Avionics Test Bed
ATM	Asynchronous Transfer Mode
ATV	Automated Transfer Vehicle
ATV-CC	ATV Control Centre
AU	Astronomical Unit
AWGN	Additive White Gaussian Noise
B2B	Business to Business
BAC	Budget At Completion
BAFA	Bundesamt für Wirtschaft und Ausfuhrkontrolle (Federal Office of Economics and Export Control)
BAPTA	Bearing And Power Transfer Assembly
BAS	Business Agreement Structure
BB	Breadboard
BCC	Battery Charge Control
BCH	Bose–Chaudhuri–Hocquenghem (a parameterized error-correcting code)
BCR	Battery Charge Regulator
BCWP	Budgeted Cost of Work Performed
BCWS	Budgeted Cost of Work Scheduled
BDC	Brushless DC (motor)
BDLI	Bundesverband der Deutschen Luft- und Raumfahrtindustrie
BDR	Battery Discharge Regulator
BEBF	Allgemeine Bestimmungen für Forschungs- und Entwicklungsverträge des Bundesministeriums für Bildung und Forschung (General Regulations for Research and Development Contracts with the German Federal Ministry of Education and Research)
BEM	Boundary Element Method
BER	Bit Error Rate
BFSK	Binary Frequency Shift Keying
BIMP	Bureau International des Poids et Mesures
Biolab	Biological Science Laboratory
BIRD	Bispectral Infra-Red Detection (satellite)
BLSS	Biological (Bioregenerative) Life Support System

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9

BNetzA	Bundes-Netz-Agentur (German Federal Network Agency)
BNR	Nonregulated Bus
BNSC	British National Space Centre
BOC	Binary Offset Coding
BOE	Basis of Estimate
BOL	Begin Of Life
BOSS	BIRD Operating System
BPSK	Binary Phase Shift Keying
BR	regulated bus
BS	Battery Simulator
BSR	semiregulated bus
BSS	Broadcasting Satellite Service
C/C	Carbon/Carbon compounds
C–SiC	Carbon–Silicon Carbide (material)
CA	Control Accounts
CAC	Cost At Completion
CAD	Computer-Aided Design
CADMOS	Centre d'aide au Développement des activités en Micropesanteur et des Opérations Spatiales (CNES, Toulouse, France)
CAIV	Cost As an Independent design Variable
CAM	Collision Avoidance Maneuver; Control Account Manager
CBS	Cost Breakdown Structure
CCB	Common Core Booster
CCD	Charge-Coupled Device
CCIR	Comité Consultatif International des Radiocommunication
CCN	Contract Change Note (Notice)
CCR	Contract Change Request
CCS	Country/Company Structure
CCSDS	Consultative Committee for Space Data Systems
CCTV	Closed Circuit Television System
CDM	Configuration and Documentation Management
CDMA	Code Division Multiple Access
CDR	Commander; Critical Design Review
CDRA	Carbon Dioxide Removal Assembly
CDTI	Centro para el Desarrollo Tecnológico Industrial
CE	Closed cycle Engine
CEAS	Confederation of European Aerospace Societies (Brussels)
CELSS	Controlled Ecological (Environmental) Life Support System
CEN	Comité Européen de Normalisation (European Committee for Standardization)
CEO	Chief Executive Officer
CEOS	Committee on Earth Observation Satellites
CER	Cost Estimation Relationship
CESS	Coarse Earth and Sun Sensor
CF	Capacity Fading
CFD	Computational Fluid Dynamics
CFRP	Carbon Fiber-Reinforced Plastics
CGP	Common Grounding Point
CHAMP	Challenging Microsatellite Payload for Geophysical Research and Application
CI	Configuration Item
CIDL	Configuration Item Data List
CL	Capacity Length; Cycle Lifetime
CM	Configuration Management; Cost Management
CMC	Ceramic Matrix Composite
CMCF	Central Monitoring and Control Facility
CMCU	Clock Monitoring and Control Unit
CMD	Command
CME	Coronal Mass Ejection
CMG	Control Momentum Gyro
CNC	Computer Numerical Control
CNES	Centre Nationale d'Etudes Spatiales
CNS	Central Nervous System
CNSA	China National Space Administration
COAS	Crew Optical Alignment Sight
COBE	Cosmic Background Explorer (NASA)
COC	Certificate Of Conformance
CoFR	Certification of Flight Readiness
CoG	Center of Gravity
Col-CC	Columbus Control Centre
COL-MU	Columbus Mock-Up
COL-TRE	Columbus Trainer Europe
COL-TRU	Columbus Trainer US
COMMS	Communication System
CON	Contractor
COP	Columbus Operations Planner
CORINE	Coordinated Information on the European Environment
COSPAR	Committee on Space Research
COSPAS	Cosmicheskaya Sistyema Poiska Avariynich Sudov (Space System for the Search of Vessels in Distress)
COTS	Commercial Off The Shelf; Components Off The Shelf
CPI	Cost Performance Index
CPLD	Complex Programmable Logic Device
CPM	Customer Product Management
CPTR	Compact Test Range
CPU	Central Processing Unit
CPV	Common Pressure Vessel
CQRM	Crew Qualification and Responsibilities Matrix
CR	Change Request; Commissioning Review
CRC	Cyclic Redundancy Code
CS	Commercial Service (Galileo)
CSA	Canadian Space Agency

CSG	Centre Spatial Guyanais (ESA launch site, Kourou, French Guiana)
CSIM	Constellation Simulator
CSP	Constraint Satisfaction Problem
CTE	Coefficient of Thermal Expansion
CU	Customer
CVCM	Collected Volatile Condensable Material
CVD	Chemical Vapor Deposition
CWS	Caution and Warning System
CWSA	Condensate Water Separator Assembly
D/C	Downconverter
D/L	Downlink
DAM	Damping Mode
DARPA	Defense Advanced Research Projects Agency (USA)
DART	Demonstrator for Autonomous Rendezvous Technology
DBS	Direct Broadcasting Satellite
DC	Direct Current
DDR	Deutsche Demokratische Republik (the former East Germany)
DDS	Data Disposition System
DEOS	Deutscher Projektvorschlag: "Orbitales Servicing" (DLR)
DET	Direct Energy Transfer
DFD	Deutsches Fernerkundungsdatenzentrum (German Remote Sensing Data Centre, DLR)
DFG	Deutsche Forschungsgemeinschaft
DFT	Data Flow Test
DGLR	Deutsche Gesellschaft für Luft- und Raumfahrt (German Society for Aeronautics and Astronautics)
DHS	Data Handling System
DIFA	Data Interface Front-end Assembly
DIFM	Direct Interface Force Method
DIN	Deutsches Institut für Normung (German Institute for Standardization)
DIODE	Détermination Immédiate d'Orbite par Doris Embarqué (Immediate On-board Orbit Determination with DORIS)
DIPS	Dynamic Isotope Power System
DLAR	Double Layer Antireflective
DLR	Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Centre)
DM	Development Model
DMS	Data Management System
DOD	Depth Of Discharge
DOF	Degrees Of Freedom
DORIS	Doppler Orbitography and Radiopositioning Integrated by Satellite
DPSK	Differential Phase Shift Keying
DR	Dynamic Range
DRD	Document Requirement Definition
DSM	Docking and Storage Module
DSMC	Direct Simulation Monte Carlo Method
DSN	Deep Space Network
DSP	Digital Signal Processor
DSPG	Distributed Single Point Grounding
DTMF	Dual-Tone Multifrequency
EAC	Estimate At Completion; European Astronaut Centre (of ESA, in Cologne)
EADS	European Aeronautic Defence and Space Company
e.c.	Economic Condition (followed by a month and year)
EC	European Commission
ECA	Evolution Cryotechnique Type A (Ariane 5 upper stage for 10 t payload)
ECD	Energy Conditioning and Distribution
ECLSS	Environmental Control and Life Support System
ECMWF	European Centre for Medium-Range Weather Forecasts
ECS	European Communication Satellite
ECSS	European Cooperation on Space Standardization
ECSS-E	ECSS E-Series Engineering Standards
ECSS-M	ECSS M-Series Management Standards
ECSS-Q	ECSS Q-Series Quality Assurance Standards
EDAC/EDC	Error Detection and Correction
EDI	External Data Interface
EDR	European Drawer Rack
EEE	Electrical, Electronic and Electromechanical
EEM	Electrical Engineering Model
EESS	Earth Exploration Satellite Service
EGNOS	European Geostationary Navigation Overlay Service
EGSE	Electrical Ground Support Equipment
EHF	Extremely High Frequency
EIDP	End Item Data Package
EIRP	Equivalent Isotropic Radiated Power
EJOP	European Joint Operations Panel
EL	Elevation
ELF	Extremely Low Frequency
ELV	Expendable Launch Vehicle
EM	Engineering Model
EMC	Electromagnetic Compatibility
EMCS	European Modular Cultivation System
EMF	Electromotive Force
EMI	Electromagnetic Interference
EML	Electromagnetic Levitator
EMU	Extravehicular Mobility Unit (space suit)
EOL	End Of Life
EOS	European On-board System
EPC	Etage Principal Cryotechnique (Ariane)

EPDS	Electrical Power Distribution System
EPIRB	Emergency Position Indicating Radio Beacon
EPM	Earth-Pointing Mode; European Physiology Module
EPROM	Erasable Programmable Read-Only Memory
EPS	Electrical Power (Sub)system
EQM	Engineering Qualification Model
ERA	European Robotic Arm (ISS)
ERIS	External Regional Integrity Service (Galileo)
ESA	European Space Agency (Paris)
ESAC	European Space Astronomy Centre (of ESA in Villafranca, Spain)
ESATAN	Thermal Software Package (of Alstom Power Aerospace)
ESC	Engineering Support Centre
ESD	Electrostatic Discharge
ESO	European Southern Observatory
ESOC	European Space Operations Centre (Darmstadt)
ESNIS	European Satellite Navigation Industries
ESRIN	European Space Research Institute (Frascati)
EST	Engineering Support Team
ESTEC	European Space Research & Technology Centre (Noordwijk)
ESTRACK	European Space Tracking and Telemetry Network
ETC	Estimate To Complete; European Transport Carrier
ETRF	European Terrestrial Reference Frame
ETS-VII	Engineering Test Satellite (NASDA)
EU	European Union
EULS	ERIS Uplink Station
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites (Darmstadt)
EUREF	Reference Frame Subcommission for Europe (IAG)
EUTEF	European Technology Exposure Facility
EUTELSAT	European Telecommunications Satellite Organization
EVA	Earned Value Analysis; Extravehicular Activity
Exp	Expander Cycle Engine
EXS	External System
FACS	Front Attitude Control System
FAE	Fixed Alkaline Electrolysis
FASTER	Facility for Absorption and Surface Tension on European Rack
FCT	Flight Control Team
FCL	Fold-back Current Limiter
FD	Flight Director; Flight Dynamics
FDF	Flight Dynamics Facility
FDIR	Fault/Failure Detection, Isolation and Recovery
FDM	Finite Difference Method
FDMA	Frequency Division Multiple Access
FDS	Flight Dynamics System
FE	Finite Element; Flight Engineer
FEC	Forward Error Correction
FEEP	Field Emission Electric Propulsion
FEM	Finite Elemente Method
FFP	Firm Fixed Price
FGB1	Russian ISS Module
FGUU	Frequency Generation and Upconversion Unit
FHG	Fraunhofer-Gesellschaft
FLPP	Future Launcher Preparatory Program
FM	Flight Model; Frequency Modulation
FMECA	Failure Mode, Effects and Criticality Analysis
FML	Fiber–Metal Laminate
FOC	Faint Object Camera; Full Operational Capability
FOG	Fiber Optic Gyro
FOV	Field Of View
FPGA	Field Programmable Gate Array
FQR	Flight Qualification Review
FREND	Front-end Robotics Enabling Near-term Demonstration
FRR	Flight Readiness Review
FS	Flight Spare; Functional Specification
FSK	Frequency Shift Keying
FSL	Fluid Science Laboratory
FSLP	First Spacelab Payload
FSS	Fixed Satellite Service
FT	Function Tree
FTF	Flygtekniska Föreningen (Swedish Society for Aeronautics and Astronautics)
FTP	File Transfer Protocol
FTS	Flight Telerobotic Servicer
GACF	Ground Assets Control Facility
GCC	Galileo Control Centre; Ground Control Center
GCR	Galactic Cosmic Ray
GCS	Ground Control Segment (Galileo)
GCS-KMF	GCS – Key Management Facility
GCT	Ground Control Team
GCTC	Gagarin Cosmonaut Training Centre
GDDN	Galileo Data Distribution Network; Global Distribution Data Network
GDV	GDV Ingenieurgesellschaft Holst mbH
GEO	Geostationary Orbit
GETEX	German ETS-VII Technology Experiments
GFW	Gesellschaft für Weltraumforschung
GGMO2	GRACE Gravity Model 02
GGTO	Galileo–GPS Time Offset
GLONASS	Globalnaja Nawigazionnaja Sputnikowaja Sistema (Global Navigation Satellite System, Russia)

GOME	Global Ozone Monitoring Experiment
GMES	Global Monitoring for Environment and Security
GMS	Ground Mission Segment (Galileo)
GMSK	Gaussian Minimum Shift Keying
GMT	Greenwich Mean Time
GNC	Guidance, Navigation and Control
GNSS	Global Navigation Satellite System
GOCE	Gravity Field and Steady-State Ocean Circulation Explorer (ESA gravity mission)
GPHS	General Purpose Heat Source
GPS	Global Positioning System (USA)
GPST	GPS Time
GQF	Gradient Furnace with Quenching
GRACE	Gravity Recovery And Climate Experiment
GRSP	Geodetic Reference Service Provider
GS	Ground Segment
GSD	Ground Sampling Distance
GSDR	Ground Segment Design Review
GSE	Ground Support Equipment
GSIR	Ground Segment Implementation Review
GSO	Geostationary Orbit
GSOC	German Space Operations Centre
GSRQR	Ground Segment Requirement Review
GSRR	Ground Segment Readiness Review
GSS	Galileo Sensor Stations
GST	Galileo System Time
G–T	Gain–Temperature relationship of an anntenna
GTO	Geostationary Transfer Orbit
GTRF	Galileo Terrestrial Reference Frame
GVT	Ground Vibration Test
H/W	Hardware
HAES	Hellenic Aeronautical Engineers Society
HAN	Hydroxylammonium Nitrate
HAP	High-Altitude Platform
HCF	High-Cycle Fatigue
HCU	Heater Control Unit
HDL	Hardware Description Language (Verilog)
HEO	Highly inclined Elliptical Orbit
HEPA	High-Efficiency Particle Filter
HF	High Frequency
HILT	Hardware In the Loop Testing
HISPASAT	Spanish communications satellite family
HLLV	Heavy-Lift Launch Vehicle
HOSC	Huntsville Operations Support Center
HPA	High-Power Amplifier
HRG	Hemispheric Resonating Gyro
HSIA	Hardware/Software Interaction Analyses
HST	Hubble Space Telescope
HTO	Horizontal Take-Off
HTP	High Test Peroxide
HTPB	Hydroxyl-Terminated Polybutadiene
HTV	H-2 Transfer Vehicle (Japanese supply vehicle for the ISS)
I/O	Input/Output (interface, module)
IABG	Industrieanlagen-Betriebsgesellschaft mbH (Munich)
IADC	Inter-Agency Space Debris Coordination Committee
IAF	International Astronautical Federation
IAG	International Association of Geodesy
IC	Integrated Circuit; Isolation and Confinement
ICBM	Intercontinental Ballistic Missile
ICD	Interface Control Document
ICDU	Integrated Control and Data Unit
ICE	Independent Cost Estimate
ICO	Intermediate-altitude Circular Orbit
ICRS	International Celestial Reference System
IDAS	Integrated Data Collection System
IDRD	Increment Description and Requirements Document
IEEE	Institute of Electrical and Electronics Engineers
IERS	International Earth Rotation Service
IF	Integrity Flag; Intermediate Frequency
IGRF	International Geomagnetic Reference Field
IGS	Interconnect Ground Subnet(work); International GNSS Service
IGSO	Inclined Geosynchronous Orbit
IIFM	Indirect Interface Force Method
IKONOS	Greek word for "imaging" – commercial Earth observation satellite
ILRS	International Laser Ranging Service
ILS	Integrated Logistic Support
IM	Integration Model
IML	International Microgravity Laboratory
IMS	Inventory Management System; Ion Mobility Spectrometer
IMU	Inertial Memory Unit
IMV	Inter-Modular Ventilation
INMARSAT	International Maritime Satellite Organization
INTELSAT	International Telecommunications Satellite Consortium
I/O	Input/Output
IODE	Issue Of Data Ephemeris
IOT	In-Orbit Testing
IOV	In-Orbit Validation
IP	International Partner
IPF	Integrity Processing Facility
IPG	Inertial Pointing Mode
IPS	Instrument Pointing System
IPV	Individual Pressure Vessel
IR	Infrared
IRAS	Infrared Astronomical Satellite
ISD	Instructional System Development

ISDN	Integrated Services Digital Network
ISL	Inter-Satellite Service
ISO	Infrared Space Observatory (ESA); International Organization for Standardization
ISP	Specific Impulse
ISPR	International Standard Payload Rack
ISRO	Indian Space Research Organization (Bangalore)
ISS	International Space Station
IST	Integrated Spacecraft Test; Integrated System Test
ISVV	Independent Software Verification and Validation
ISY	International Space Year
IT	Information Technology
ITAR	International Traffic in Arms Regulations
ITCB	International Training Control Board
ITRF	International Terrestrial Reference Frame
ITRS	International Terrestrial Reference System
ITU	International Telecommunications Union (United Nations, Geneva)
JAXA	Japan Aerospace Exploration Agency
JEM	Japanese Experiment Module (ISS)
JGM	Joint Gravity Model
JOS	Japanese On-board System
JPL	Jet Propulsion Laboratory (NASA, Pasadena, California)
JSC	Johnson Space Center (NASA, Houston, Texas)
KIP	Key Inspection Point
KMF	Key Management Facility
KOS	Keep Out Sphere
KSC	Kennedy Space Center (NASA, Cape Canaveral, Florida)
L1, L2	Lagrange points
LAM	Large-Angle Maneuver
LAN	Local Area Network
LBA	Luftfahrt-Bundesamt
LCF	Low-Cycle Fatigue
LCL	Latching Current Limiter
LCOS	Liquid Carry-Over Sensor
LCT	Laser Communication Terminal
LDAP	Lightweight Directory Access Protocol
LED	Light-Emitting Diode
LEO	Low Earth Orbit
LEOP	Launch and Early Orbit Phase
LET	Linear Energy Transfer
LF	Loop Frequency; Low Frequency
LGF	Low-Gradient Furnace
LH_2	Liquid Hydrogen
LHCP	Left Hand Circular Polarization
LHP	Loop Heat Pipe
LIB	Lithium-Ion Battery
LISA	Laser Interferometer Space Antenna
LLI	Long Lead Item
LM	Link Margin
L-MSS	Land Mobile Satellite Service
LNA	Low-Noise Amplifier
LOC	Lines Of Code
LORAN	Long-Range Navigation
LOS	Line of Sight; Loss of Signal
LOX	Liquid Oxygen, LO_2
LRR	Launch Readiness Review
LRR	Laser Ranging Reflector
LSA	Logistic Support Analysis
LSS	Life Support System
LT	Low Temperature
LTL	Low-Temperature Loop
LUT	Local User Terminal
M&C System	Management & Control System
MAIT	Manufacturing, Assembly, Integration, Testing
MARS	Microgravity Advanced Research and Support Center
MC	Magnetic Cleanliness
MCA	Multiconstituent Analyzer
MCC	Mission Control Center
MCC-H	Mission Control Center Houston
MCC-M	Mission Control Centre Moskau
MCDR	Mission Critical Design Review
MCOP	Multilateral Crew Operations Panel
MCRR	Mission Commissioning Results Review
MCS	Master Control Station; Monitoring and Control System
MCU	Micro-Control Unit
MDDN	Mission Data Dissemination Network
MDR	Mission Definition Review
MECO	Mechanical Configuration
MEO	Medium-Altitude Earth Orbit
MER	Mars Exploration Rover
MERIS	Medium-Resolution Imaging Spectrometer
MET	Mission Elapsed Time
MF	Medium Frequency
MFSA	Magnetic Field Simulation Assembly
MGF	Message Generation Facility
MGSE	Mechanical Ground Support Equipment
MHD	Magnetohydrodynamics
MIL	Military Standard
MIP	Mandatory Inspection Point
MIR	Medium-wave Infrared; Russian word for "peace" – former Russian space station
MKMF	Mission Key Management Facility
MLI	Multilayer Insulation
MLM	Russian module on the ISS
MM	Massemodell

MMC	Metal Matrix Composite
MMH	Monomethylhydrazine
MMI	Man–Machine Interface
M-MSS	Maritime Mobile Satellite Service
MMU	Manned Maneuvering Unit; Memory Management Unit
MoI	Moment of Inertia
MOMS	Modular Optoelectronic Multispectral Scanner
MON	Mixed Oxides of Nitrogen
MORABA	Mobile Rocket Base
MoS	Margin of Safety
MPD	Magnetoplasmadynamic (thruster)
MPG	Max-Planck-Gesellschaft
MPLM	Multipurpose Logistics Module
MPP	Maximum Power Point
MPPT	Maximum Power Point Tracking
MPS	Mission Planning System
MRT	Mission Readiness Test
MSAS	Multifunctional Satellite Augmentation System (Japan)
MSC	Monitoring and Control Center
MSDR	Mission System Definition Review
MSF	Mission Support Facility
MSFC	Marshall Space Flight Center (NASA, Huntsville, Alabama)
MSG	Meteosat Second Generation; Microgravity Science Glovebox
MSK	Minimum Shift Keying
MSL	Materials Science Laboratory
MSS	Mobile Satellite Service
MSU	Monitoring and Safety Unit
MT	Moderate Temperature
MTCR	Missile Technology Control Regime
MTFR	Mean Time For Repair
MU	Mock-Up
MUCF	Mission and Uplink Control Facility
MUSC	Microgravity User Support Centre (of DLR, in Cologne)
NAG	Naval Astronautics Group
NASA	National Aeronautics and Space Administration (USA)
NASA-STD	NASA Standard
NASDA	National Space Development Agency of Japan (now JAXA)
NASTRAN	NASA Structural Analysis System (FEM software)
NBF	Neutral Buoyancy Facility
NC	Normally Closed
NCR	Nonconformance Report
NEA	Nonexplosive Actuator
NGSO	Nongeostationary Orbit
NHB	NASA Handbook (Quality and Safety)
NIMA	National Imagery and Mapping Agency (USA)
NIR	Near Infrared
NO	Normally Open
NOAA	National Oceanic and Atmospheric Administration (USA)
NOM	Nominal
NORAD	North American Aerospace Defense Command (now USSTRATCOM)
NPL	National Physical Laboratory (UK)
NPSHR	Net Positive Suction Head Required (relative pump suction height)
NRZ-L	Nonreturn to Zero Level
NRZ-M	Nonreturn to Zero Mark
NRZ-S	Nonreturn to Zero Space
NSGU	Navigation Signal Generator Unit
NSTS	National Space Transportation System (NASA)
NTO	Nitrogen Tetroxide
NTP	Network Time Protocol
NVvL	Nederlandse Vereniging voor Luchtvaart-techniek
O	Operator
OASPL	Overall Sound Pressure Level
OBC	On-Board Computer
OBDH	On-Board Data Handling
OBP	On-Board Processor
OBS	On-Board Switching
OBSM	On-Board Software Maintenance
OC	Operations Coordinator
OCS	Office Communication System; Orbit Control System
OCT	Orbit Control Thruster
OD&TS	Orbit Determination & Time Synchronization
OE	Open Cycle Engine
OGC	Office of Global Communication (USA)
OGSE	Optical Ground Support Equipment
OHA	Operating Hazard Analyses
OHB	Orbitale Hochtechnologie Bremen
OLEV	Orbital Life Extension Vehicle
OMS	Orbit Maneuvering System
OOK	On–Off Keying
OOS	On-Orbit Servicing; On-Orbit Summary
OPF	Operation Preparation Facility
OPS	Operations
OQPSK	Offset Quad-Phase Shift Keying
ORC	Organic Rankine Process
ORDEM	Orbital Debris Environmental Model
ORR	Operational Readiness Review
ORU	Orbital Replacement Unit
OS	Open Service (Galileo)
OSPF	Orbit and Synchronization Processing Facility
OSR	Optical Solar Reflector

OSTC	On-Stage Thermal Cycling
OSTP	On-orbit Short-Term Plan
OTS	Off The Shelf
P	Prime (contractor, investigator)
P/L	Payload
PA	Product Assurance
PA&S	Product Assurance and Safety
PAD	Parts Approval Documents
PAF	Processing and Archiving Facility
PBS	Project Breakdown Structure
PC	Personal Computer
PCA	Pressure Control Assembly
PCDF	Protein Crystallization Diagnostic Facility
PCDU	Power Control and Distribution Unit
PCM	Pulse Code Modulation
PCS	Portable Crew Station
PDE	Propulsion Drive Electronics
PDR	Preliminary Definition Review; Preliminary Design Review
PE	Pressure-Fed Engine
PEEK	Polyetheretherketone
PEM	Proton Exchange Membrane
PEO	Polar Earth Orbit
PER	Packet Error Rate
PFM	Prototype Flight Model
PHA	Preliminary Hazard Analyses
PHM	Passive Hydrogen Maser
PI	Principal Investigator
PIA	Propellant Isolation Assembly
PID	Proportional, Integral and Differential (controller)
PIM	Passive Intermodulation Products
PKMF	PRS Key Management Facility
PL	Path Loss
PLL	Phase-Locked Loop
PLT	Pilot
PM	Phase Modulation; Pressurized Module; Project Management
PMD	Propellant Management Device
PMP	Parts, Materials and Processes; Project Management Plan
PNT	Position–Navigation–Time
POCC	Payload Operations Control Center
PoI	Products of Inertia
POIC	Payload Operations and Integration Center
POSEIDON	Positioning Ocean Solid Earth Ice Dynamics Orbiting Navigator
P-POD	Poly Picosatellite Orbital Deployer
PPM	Parts Per Million
PPP	Public–Private Partnership
PPT	Pulsed Plasma Thruster
PRF	Pulse Repetition Frequency
PRN	Pseudo-Random Noise
PRR	Preliminary Requirements Review; Propellant Refillable Reservoir
PRS	Public Regulated Service
PSD	Power Spectral Density
PSK	Phase Shift Keying
PSLV	Polar Satellite Launch Vehicle (India)
PSR	Preshipment Review
PSS	Procedures, Standards and Specifications (ESA)
PT	Product Tree; Project Team
PTB	Physikalisch-Technische Bundesanstalt
PTC	Positive Temperature Coefficient
PTF	Precise Time Facility
PTFE	Polytetrafluorethylene
PUS	Packet Utilization Standard
PV	Pyrotechnic Valve
PVA	Photovoltaic Array
PVD	Physical Vapor Deposition
PVE	Photovoltaic Energy
PVT	Pressure–Volume–Temperature
PWM	Pulse Width Modulation
PWS	Portable Workstation
QA	Quality Assurance
QM	Qualification Model; Quality Management
QPSK	Quadrature Phase Shift Keying
QR	Qualification Review
QRB	Qualification Review Board
QSP	Quality Structure Plan
R&D	Research and Development
RAAN	Right Ascension of the Ascending Node
RAeS	The Royal Aeronautical Society
RAFS	Rubidium Atomic Frequency Standard
RAM	Random Access Memory (Read/Write)
RAMS	Reliability, Availability, Maintainability, Safety
RBE	Rigid Beam Element
RC	Resistance – Capacitor (lowpass)
RCS	Reaction Control System
RCV	Receiver
RF	Radiofrequency
RFA	Request For Approval
RFI	Request For Information
RFP	Request For Proposal
RFQ	Request For Quotation
RFW	Request For Waiver
RHCP	Right Hand Circular Polarization
RHU	Radioactive Heater Unit
RID	Review Identified Discrepancy
RIT	Radiofrequency Ion Thruster
RLG	Ring Laser Gyros
RLV	Reusable Launch Vehicle
RM	Risk Management

RML	Recovered Mass Loss
RMS	Root Mean Square
RNG	Ranging
RNRZ	Randomized Nonreturn to Zero
RNSS	Radio Navigation Satellite Service
ROD	Review Of Design
ROKVISS	Robot Component Verification on the International Space Station
ROM	Read-Only Memory
ROS	Russian On-orbit Segment; Russian On-board System
ROTEX	Robot Technology Experiment
RR	Radio Regulations; Requirements Review
RS	Reed–Solomon (convolutional code)
RSA	Russian Space Agency
RST	Raumfahrt- und Systemtechnik GmbH (Warnemünde)
RTG	Radioisotopic Thermoelectric Generator
RTK	Real Time Kinematik
RTM	Radio link Test Model; Resin Transfer Molding
RVD	Rendezvous and Docking
S	Specialist
S/C	Spacecraft
S/W	Software
SA	Solar Array
SAA	South Atlantic Anomaly
SADM	Solar Array Drive Mechanism
SAR	Search And Rescue; Solar Array Regulator; Synthetic Aperture Radar
SARSAT	Search And Rescue Satellite-Aided Tracking
SAS	Space Adaption Syndrome
SCA	Solar Cell Assembly; Système de Contrôle d'Attitude (Ariane)
SCCF	Spacecraft and Constellation Control Facility
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Cartography
SCOE	Special Checkout Equipment
SCOS	Spacecraft Operating System
SCPF	Satellite (Spacecraft) Constellation Planning Facility
SCR	Solar Cosmic Radiation
SCUBA	Self-Contained Underwater Breathing Apparatus
SD	Solardynamik
SDDN	Satellite Data Distribution Network
SDHS	Site Data Handling Set
SE	Systems Engineering
S-EDDN	Satellite External Data Distribution Network
SEE	Single Event Effect
SEL	Single Event Latchup
SES	Societé Européenne des Satellites (Luxembourg) (ASTRA)
SEU	Single Event Upset
SFDU	Standard Format Data Unit
SFOG	Solid Fuel Oxygen Generator
SFS	Standard Frequency and Time Signal Satellite Service
SG	Solar Generator
SGP	Simplified General Perturbation
SHA	Subsystem/System Hazard Analyses
SHF	Super High Frequency
SHOGUN	Shock Generating Unit
SIM	Simulator
SIMIS	Simulation Mission Study Group
SIS	Signal In Space
SISA	Signal In Space Accuracy
SISMA	Signal In Space Monitoring Accuracy
SLA	Service Level Agreement
SLR	Satellite Laser Ranging
SM	Russian module on the ISS; Structural Model
SMS	Satellite Media (Message) Services
SNAP	System for Nucelar Auxiliary Power
SOC	State Of Charge; System on One Chip
SoL	Safety of Life
SOS	Space Operation Service
SOW	Statement Of Work
SPDM	Special Purpose Dexterous Manipulator
SPE	Solar Particle Event
SPF	Service Products Facility; Sun-Pointing Fixed Mode
SPI	Schedule Performance Index
SPL	Sound Pressure Level
SPM	Suspend Mode
SPOT	Systeme Pour l'Observation de la Terre
SPT	Stationary Plasma Thruster
SPV	Single Pressure Vessel
SR	Shunt Regulator; Space Research Service
SRAM	Static Random Access Memory
SRM	Solid Rocket Motor
SRMS	Shuttle Remote Manipulator System
SRR	System Requirements Review
SRS	Shock Response Spectrum; Space Research Service
SS	Summer Solstice
SSC	Swedish Space Corporation
SSCC	Space Station Control Center
SSDS	Self-Supplied Diving System
SSIPC	Space Station Integration and Promotion Center (JAXA, Tsukuba, Japan)
SSM	Second Surface Mirror
SSME	Space Shuttle Main Engine
SSO	Sun-Synchronous Orbit
SSP	Space Station Program (Quality and Safety)
SSPA	Solid-State Power Amplifier
SSPC	Solid-State Power Controller
SSRMS	Space Station Remote Manipulating System

SST	Sea Surface Temperature
SSTF	Space Station Training Facility
SSTL	Surrey Satellite Technology Ltd (UK)
SSTO	Single Stage To Orbit
STM	Structure and Thermal Model
STS	Space Transportation System (the Space Shuttle)
SUMO	Spacecraft for the Universal Modification of Orbits (DARPA)
SURV	Survival/Safe (limits)
SVF	Software Validation Facility
SVFW	Schweizerische Vereinigung für Flugwissenschaften (Swiss Association of Aeronautical Sciences)
SVT	System Validation Test
SW	Software
SWT	Science Working Team
TAI	Temps Atomique International (International Atomic Time)
TB	Thermal Balance
TC	Telecommand; Thermal Cycling
TCA	Thrust Chamber Assembly
TCP/IP	Transmission Control Protocol/ Internet Protocol
TCPI	To-Complete Performance Index
TCS	Thermal Control Subsystem; Trajectory Control Sensor
TCV	Temperature Control Valve
TD	Thermal Distortion
TDMA	Time Division Multiple Access
TDRSS	Tracking and Data Relay Satellite System
TECSAS	Technology Satellite for Demonstration and Verification of Space Systems
TEG	Thermoelectric Generator
TEMPUS	Tiegelfreies Elektro-Magnetisches Prozessieren Unter Schwerelosigkeit (Containerless Electromagnetic Processing Under Weightlessness)
TEXUS	Technologische Experimente Unter Schwerelosigkeit (Technology Experiments Under Microgravity)
TID	Totally Ionizing Dose
TIR	Thermal Infrared
TLE	Two-Line Element
TLM	Telemetry
TM	Telemetry; Thematic Mapper (Landsat sensor); Thermal Model
TM/TC	Telemetry/Telecommand (subsystem)
TMF	Thruster Management Function
TML	Total Mass Loss
TMR	Triple-Module Redundancy
TNC	Terminal Node Controller
TOPEX	Ocean Topography Experiment
TP	Tactical Plan
TQM	Total Quality Management
TQVS	Training, Qualification and Verification System
TR	Transmit and Receive
TRL	Technology Readiness Level
TRR	Test Readiness Review; Training Readiness Review
TS	Technical Specification
TSP	Time Service Provider
TTC/TT&C	Telemetry, Tracking and Command
TTCF	Telemetry,Tracking and Control Facilities
TVC	Thermal Vacuum Cycling
TWSTFT	Two-Way Satellite Time and Frequency Transfer
TWT	Traveling Wave Tube
U	User
U/C	Upconverter
U/L	Uplink
UDMH	Unsymmetrical Dimethylhydrazine
UHF	Ultrahigh Frequency
UHV	Ultrahigh Vacuum
ULS	Uplink Station
UNCOPUOS	United Nations Committee on the Peaceful Uses of Outer Space
UNIDROIT	Institut International pour l'Unification du Droit (International Institute for the Unification of Private Law)
UPS	Unified Propulsion System
USOC	User Support and Operations Centre
USOS	US On-Board System; US On-Orbit Segment
USSPACECOM	US Space Command (Colorado, USA)
UTC	Coordinated Universal Time
UV	Ultraviolet
V/T Method	Voltage–Temperature Method
VAC	Variance At Completion
VCD	Verification Control Document
VDI	Verband deutscher Ingenieure
VHF	Very High Frequency
ViAS	Video Archive System
VIS	Visible Light (spectral range)
VLAN	Virtual Local Area Network
VLBI	Very Long Baseline Interferometry
VLF	Very Low Frequency
VOA	Volatile Organic Analyzer
VOC	Volatile Organic Compound
VoCS	Voice Communication System
VOSDUKH	Russian carbon dioxide removal equipment on the ISS
VPN	Virtual Private Network
VSOC	Venus Express Science Operations Center

VT	Voltage–Temperature
VTO	Vertical Take-Off
WAAS	Wide Area Augmentation System
WAOSS	Wide Angle Optoelectronic Stereo Scanner
WBS	Work Breakdown Structure
WDE	Wheel Drive Electronics
WGS	World Geodetic System
WLP	Weekly Look-ahead Plan
WMO	World Meteorological Organization
WP	Work Package
WPD	Work Package Description
WRC	World Radiocommunication Conference
WS	Winter Solstice
WTA	Warning Time Analyses
WVR	Water Vapor Regained
XDA	X-band Downlink Assembly
XMM	X-ray Multimirror
XPD	Cross-Polarization Discrimination
ZARM	Zentrum für Angewandte Raumfahrttechnologie und Mikrogravitation (Centre of Applied Space Technology and Microgravity, Bremen)
ZUP	Zentr Upravlenija Poljotami (Flight Control Centre, Moscow)

Symbol List

a	acceleration; average orbit radius, satellite; relative frequency drift; semi-major axis; speed of sound; transition loss
a_{ij}	Runga–Kutta coefficients
a_{MAX}	maximum acceleration of a signal
a_{RMS}	average acceleration of a signal
A	aperture; area, cross-section (surface area); azimuth
$\boldsymbol{A}$	information matrix
A, a	critical exponents
A_{a}	area of nozzle exit
A_{b}	combustion front surface
AA	additional attenuation
AD	atmospheric attenuation
b	receiver clock bias; semi-minor axis
b_i	Runge–Kutta coefficients
B	bandwidth; Earth's magnetic flux density; magnetic field
B_{s}	static magnetic field
c	exhaust velocity; specific heat; speed of light
$\overline{c}$	average thermal velocity
c_0	speed of light in vacuum
c^*	characteristic speed
c_i	mass fraction of species i; Runge–Kutta coefficients
c_{P}	specific heat at constant pressure
c_{V}	specific heat at constant volume
C	carrier power
C_{D}	drag coefficient
C_{F}	thrust coefficient
C_{l}	lift coefficient
C_{nm}	spherical coefficient
C_{p}	aerodynamic pressure coefficient
C_{R}	radiation pressure coefficient
d	cable attenuation; thickness, distance
d_{s}	specific diameter
D	antenna diameter; attenuation constant; Damköhler number; diameter; diffusion coefficient; drift; radiation dose
$\boldsymbol{D}$	dipole vector
D_{e}	equivalent dose
$\boldsymbol{e}$	eccentricity vector
e	specific energy
$e_{\mathrm{A,IR,S}}$	visibility factor
$\boldsymbol{e}_{\mathrm{Sun}}$	unit vector, Earth–Sun
$\boldsymbol{e}_{x,\,y,\,z}$	unit vectors
E	eccentricity anomaly; energy
$\boldsymbol{E}$	transformation matrix
E_0	starting value
E_{i}	anomaly
E_{radiat}	energy of radiation
$\boldsymbol{f}$	vector function
f	frequency
f_{Cycle}	cycle lifetime
f_{P}	plasma frequency
f_{T}	carrier frequency
F	area; gravitational force; particle flux; thrust
F_1, F_2	focus
F_{B}	magnetic force
FD, fd	free space dispersion
F_{g}	Earth attraction force
F_{g}	gravitational force
F_{L}	Lorentz force
F_{N}	nominal force
F_{R}	friction force
F_{f}	fictitious force
F_{T}	inertial force
F_{z}	centrifugal force; centripetal force
g	gravitational acceleration (Earth); model parameter
g_0	gravitational acceleration at the Earth's surface
G	antenna gain; gravitational constant

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9

h	altitude; mathematical model of measurement; Planck constant; step size
h_{ct}	contact with heat conduction
H	angular momentum; altitude
$\boldsymbol{H}$	Jacobian matrix
i	inclination/orbital inclination
$\boldsymbol{i}$	inclination vector
I	impulse; electric current; moment of inertia; power
I_0	current
I_{sp}	specific impulse
j_c	mass flow
J	carrier frequency; cost function
J_n, J_{nm}	coefficients of gravity field
J_{xx}–J_{zz}	moments of inertia
k	Boltzmann constant; form factor; load factor
K	heat capacity; Kalman gain; nodes
$\boldsymbol{K}$	stiffness matrix
Kn	Knudsen number
K_p	geomagnetic index
$K_{p,\,i,\,d}$	parameters (proportional, integral, differential) of a PID controller
l	mean longitude
L	lift; loss; typical system length; heat conduction
L_k	capillary length
m	mass (space vehicle, satellite); particle mass
$\dot{m}$	mass flow
m_r	modal mass
M	Earth mass; Mach number; mass of the central body; mean anomaly, average angle; molecular weight momentum
$\boldsymbol{M}$	mass matrix
M_D	angular momentum vector
M_{grav}	gravity momentum
M_i	dipole moment
M_s	mass of a perturbing body
n	angular velocity; number of impacts
n_e	electron density
n_{opt}	optimum number of rocket stages
n_r	normalization factor
n_s	specific rotational speed
N	mean thermal noise; number of expected particles
N_E	number of eclipses
Nu	Nusselt constant
p	impulse; orbit parameters; pressure; probability; signal availability
$\boldsymbol{p}$	force model parameter (vector); vector of external force
p_a	nozzle pressure
p_{dyn}	dynamic pressure
p_{stat}	static pressure
p_{t2}	Pitot pressure
P	antenna diagram; power; primary bus power; transmission power
$\boldsymbol{P}$	covariance matrix
P_0	pressure
P_{nm}	Legendre polynomial of first degree
Pr	Prandtl constant
q	electrical charge; quaternions; rotation increment; thermal flow
Q	integral heat; process noise
r	geocentric satellite position; orbit radius
$\boldsymbol{r}$	position vector
$\dot{r}$	velocity vector
$\ddot{r}$	acceleration vector
r_i	distance from test object
r_s	geocentric position vector of disturbing body
R	coordinates of ground station; distance of centers of mass; gas constant; maximum error free data rate; radius; resistance; respiratory index; station coordinates
R_0	radius of the probe
Ra	Rayleigh constant
RC	time constant
Re	Reynolds number
R_S	actual distance, Earth–Sun; radius of the Sun
$\boldsymbol{R}_{x,y,z}$	matrices

s	position coordinates
$\boldsymbol{s}$	topocentric position vector
s, s_E, s_N, s_Z	topocentric satellite position; east, north, zenith
S	entropy; reference area; Soret coefficient; signal power; solar flux density; surface area
$\boldsymbol{S}$	sensitivity matrix
S/N	signal-to-noise ratio
Sc	Schmidt constant
SC	solar constant
S_{nm}	gravitational field coefficient (potential coefficient); harmonic coefficients
S_s	solar energy flux density
S_s^{total}	total energy flux
St	Stanton constant
t	epoch; time
t_0	start time
t_{ae}	eclipse duration
$t_{rec,sys,sky}$	system noise temperature
T	radiation temperature; orbital period; temperature; torque
T_b	temperature of combustion chamber
T_c	critical temperature
T_{eff}	effective radiation temperature
T_H	background temperature
T_{WG}	wall temperature
u	argument of latitude; circular velocity; controller exit (time domain)
u_a	gas velocity
$\boldsymbol{U}$	time-dependant transformation from inertial to Earth-coupled system
U_0	voltage
U_{oc}	voltage at working point
v	velocity
v_R	average orbital velocity
$\boldsymbol{v}_{rel}$	velocity vector
V	potential; velocity; volume
V_{ab}	combustion front velocity
w	mass flow density
$\boldsymbol{W}$	weight matrix
$\boldsymbol{x}$	vector of estimation parameters; vector of translation
x, x_{ef}	coordinates in the equatorial plane
x, y, z	Cartesian coordinates
$\hat{X}$, $\hat{Y}$	Cartesian coordinates within the orbital plane
$x_{R,B}$	vector in reference, body system
$y(t)$	state vector
y, y_{ef}	coordinates within the equatorial plane
z	measurement; scalar measurement
$\boldsymbol{z}$	measurement vector
z, z_{ef}	Earth axis
Z	compressibility factor
Υ	Vernal equinox
∂	partial differentiation
∇	gradient
α	reference axis of ellipsoid; right ascension; rotation angle; solar absorption
α_λ	absorbed dose (wavelength)
β	modulation index; reflectivity; thermal extension coefficient
γ	adiabatic exponent; gravity constant; spectral noise density; surface tension
Γ	attenuation
δ	surface/flow angle
Δ	delta change; difference
ε	attitude deviation signal; elevation angle; emissivity; reflectivity
ε_λ	emitted radiation as a function of wavelength
η	efficiency; viscosity
Θ	astronomical time; azimuth; orbit angle
κ	heat of evaporation; thermal diffusivity

λ	geographic longitude; mean free path length; thermal conductivity; wavelength of radiation
μ	dynamic viscosity; friction
μ_0	dipole moment; magnetic permeability
ν	kinematic viscosity; true anomaly
ρ	density
ρ_c	critical density
σ	electrical conductivity; energy flux density; standard deviation; Stefan–Boltzmann constant; surface tension
τ	dimensionless temperature
φ	attitude angle; geographic latitude; opening angle; phase shift
$\boldsymbol{\Phi}$	elevation; radiation flux
$\boldsymbol{\Phi}$	transfer matrix
χ	susceptibility
Ψ	pressure parameter
Ψ_r	vibration shape vector
ω	angular velocity; argument of perigee; modulation frequency
ω_r	eigenfrequency
Ω	right ascension of the ascending node

Index

3D displays and animations 479
7th Framework Program 833
850 W of end-of-life DC power 721

A
ABAQUS 218
ability to radiate 271
ablation cooling 89
abrasion 231
absolute attitude 341
absorbed environmental energy 269
absorptance 271
absorption of solar energy 290
acceleration 327
acceleration rockets 160
acceptable quality 198
acceptance data package (ADP) 755
acceptance level 223, 675
Acceptance Planning 755
acceptance problems 242
Acceptance review (AR) 750
Acceptance test 192, 282, 665, 669
access port 702
acknowledgment 472
acoustic energy 675
acoustic load spectra 204, 223
acoustic noise test 674
acoustic spectrum 674
Acoustic Tests 674
acquisition of the satellite signal 505
activated carbon cartridge 431
activated carbon and LiOH cartridges 426
active electronic current limiter 266
Active pixel sensors 343
actively cooled 512
actively fed propellant 307
actuators 149, 351
Adams–Bashforth method 65
adaptive interface 229
adaptive structural elements 228
adhesion, friction 231
ADN (Ammonium Dinitramide) 307
advanced very high-resolution radiometer 523
aerobraking 48, 591
aerodynamics 82
aerogels 588
aerothermodynamics 33
age of the Universe 577
aggressive propellants 321
Agility 333
agreements with other launching states 823
Air-breathing concepts 136
air-conditioning 422
air drag 449
Air drag and steering losses 120
Air exchange 430
air quality control system 430
air revitalization system 424
airbag system 589
airborne simulator 716
airlock 431
Al_2O_3 granular material 316
albedo 37, 269
albedo exclusion angle 343
A-level 223
algae reactors 439
alkaline electrolysis 424
all-in-one concept 184
all-titanium tanks 309
all-weather capabilities 528
altimeters 589
aluminum 288
aluminum alloys 205
ambient pressure thermal cycling (APTC) test 684
ammonia 288, 432
amplitude of the vibration load 672
analog film 528
analog sensor 734

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann

© 2009 John Wiley & Sons, Ltd.

ISBN: 978-0-470-69739-9

Analogy 801
analogy to electrical laws 278
Analysis validation test 665
Analytical Orbit Models 61
analytical solution 446
angular deviation 334
angular differences 341
angular momentum 337
angular velocity 54
anomaly report 470
ANSYS 218
antenna characteristic 729
antenna control system 650
antenna feed 650
Antenna Pointing 498
Antenna Tracking 489
antennas 21, 573, 693
AOCS (Attitude and Orbit Control System) 292
aphelion 587
apogee boost maneuvers 27
apogee engines 304
Apogee Injection 301, 303
Apogee thrusters 160
Apollo 11, 12, 14, 15, 16, 17, 580
approach ellipsoid 449
approach strategy 447
approach velocity 450
appropriate quality standards 785
approximate starting value 70
a priori value 70, 73
aramid fibers 206
arc-heated test facilities 96
architecture of the propellant tanks 149
architecture of the space mission 647
archiving and cataloging system 534
arcing 46
Arcjets 318
arcjet thruster 319
argument of perigee 54
Ariane 7
Ariane 5 130, 721
Arianespace 117
Arithmetic nodes 279
array antennas 531
arrival velocity 587
ascending node 54
ASKA 218
Assembly, Integration and Test 704
assembly, integration and validation 754
assessment study 462
ASTRINE 424
astronauts with special responsibilities 403
astronomical objects 27
Atlas V 131
atmosphere 529, 582
atmosphere control and supply system 419
atmosphere revitalization system 423
Atmospheric Attenuation 500
atmospheric density 59
atmospheric drag 47
atmospheric friction 522
Atmospheric Influences 47
atomic clocks 555, 558
atomic oxygen 34, 211, 212
atomic oxygen ratio 47
attestation 776
attitude 334
attitude and orbit control system 733
attitude control 302, 303, 332, 340, 451
Attitude Control System 332, 357
Attitude control thrusters 160, 455
Attitude determination 332, 350
attitude deviation 341
attitude dynamics 337
attitude kinematics 337
attitude prediction 340
attitude propagation 340
Attitude Sensors 341
attitude sensors and actuators 333
attitude status 341
attribute the risk of a launch failure 827
ATV 14, 443, 514
Auditing of suppliers 769
authorization of space activities 818
automatic planner 481
automatic systems 443
Autonomous sending 478
auxiliary power supply 248
Availability 769
availability of a space segment and a ground segment 777
aviation and maritime search and rescue 718
Avionics and Energy Supply 146
Avionics Test Bed (ATB) 733
AX.25 Protocol 701, 706
axial grooves 288
Axial turbines 176
axially oriented thrusters 302

B

back rooms 475
background noise 48
background radiation 46
bacteria and odor filter 429
BAFA 821
baffle 342
bake-out 50
bakeout test 683
baking 50
balance of calories, vitamins and minerals 434
balanced thermal conditions 721
ballistic coefficient 91
ballistic flight phase 141
bandwidth 532
barometric altitude formula 594
barometric pressure 49
baseband 734
baseband equipment 488, 491
baseline training 508
battery 257
battery management system 261
BDC motor 228
beacon signal 701
beam elements 217
beam-forming networks 729
bearing and power transfer assembly 216
bed loading 318
Belleville spring washer 226
benchmarking 771, 797
BER 503
beryllium alloys 205
best effort 827
Beta cloth 290
bimetallic mechanism 290
binary data 701
biofilters 439
biological components 439
biological contamination 442
biological starter cultures 440
biomass 524
Biorhythms 615
Biosphere-2 439
biosphere 418
BIOS Program 439
Bipropellant 303
BIRD (Bispectral Infrared Detection) 707
bit error rate 494
bit flips 362
Bi-Φ-L code 392
Black paint 272
bladder tank 311
blind inserts 212
Blow-Down Mode 307
blow-down ratio 308
blow tanks 427
body-mounted arrays 255
body system 334
body with an atmosphere 589
Boltzmann equation 98
bonding 732
bonding conditions 734
booms 216
booster 136, 155
Boost stages 121
Bottom-up Cost Estimate 801
boundary element method 219
Boundary nodes 279
box architecture 710
Braking maneuvers 302
braking rockets 589
Brayton (Joule) process 240
breadboard models 204, 222
breadboards 703
breathing oxygen 424
Broadcast architecture 650
broadcasts 540
buck and boost regulators 266
buoyancy 594
Business Agreement Structure 799
business-to-business (B2B) procurement 808
bus voltage 723

C

cablecutter 223
CAD model 218
calculate satellite orbits 525
calendar life 258
calibration 468, 734
capability 818
capillary forces 423
capillary pressure 287
capillary structures 288
capillary tube 317
carbon/carbon compounds 206
carbon fiber-reinforced plastic 206

carbon fibers 206
Cardiovascular System 611, 617
carrier frequencies 555, 570
carrier levels 503
carrier power 502
carrier-to-interference ratio 501
cartridges 431
catalog entry 472
catalyst beds 317
Catalysts 315, 318
catalytic surface 87
catastrophic failures 435
Cause of Risk 759
caution and warning system 508
cavitation 143
cavity emitter 36
CCD chip 343
CCDs 48
CDMA 555
CDRA 424
ceiling outlets (diffusors) 422
cell balancing 260
center of gravity 137
central control unit 153
central monitoring and control facility (CMCF) 563
central unit for controlling and monitoring 723
centrifugal force 592
centrifugal separator 423
ceramic matrix composites 206
Certification and qualification 769
certified and implemented 768
CFRP laminate 228
CFRP tube insert 213
Change in Electrical Resistance 47
change in inclination 77
Change management 791
characteristic areas of application 26
charge regulator 260
Checking its processes 769
checks the manufacturing 774
chemical and electrical propulsion 700
chemical purity 305, 313
chemical reactions 86
chlorella algae reactor 439
choice of the applied material 153
chugging 179
circuit design effort 265
circular orbit 447
Class 1 Catastrophic 739
Class 2 Critical 739
Class 3 Considerable 740
Class 4 Negligible 740
clean pad concept 185
clean room 185, 213
clean-room class 663
clean-room conditions 704
Cleaning 213
cleanliness 152
climate 184
climate effects 520
climatology 529
clinical relevance 620
clock accuracy 736
clock monitoring and control unit 558
clock signals 726
clock technology 554
closed control loop 332
closing phase 445
closing the loops 434
Clothing 432
CNC tools 209
Coarse Sun sensors 344
Coaxial Injection 164
COBE 577
coding 551
cold background 34, 705
cold backup 478
Cold Gas 302
cold gas systems 307
cold-welding 232
collect space debris particles 102
collected volatile condensable material 211
collectors 240
collimated radiation 36
collimation mirror 688
collision avoidance maneuver 450
collision risk 103
colocating 81
Columbus 423
Columbus Control Center 505
Columbus laboratory module 13
Columbus module 419, 505
Columbus program 15
combination of H_2/O_2 156
combiner 492
combustion chamber 317, 321

combustion velocity 162
Comet Halley 585
comfort range 421
command requests 472
command test 505
commanding 465, 471
commands to the thrusters 448
commercial off-the-shelf (COTS) 225, 699, 758, 779
commercial operation 27
commercial procurement approach 747
commercial service (CS) 571
commercial space activities 1
commissioning and in-orbit test 468
commissioning phase 659
Commitment 744
common bulkhead 145
common language 406
common panels 510
communication architecture 649
communication and navigation 520
Communication and Reporting 757
communication link 445
communication officer 512
communication 483, 508, 584
communications satellites 17, 485, 547
compact satellite 710
Compact Test Range 693
company manual 772
COMPASS 557
compatibility with the wall material 287
compatible with the transmissions of other RNSS systems 718
complete GPS constellation 561
complete system 22
Complex distribution services 535
complex signals 734
complexity 305, 657, 760
composite tanks 309
composition of an ISS expedition crew 405
composition of each crew 402
compressed data 533
Compression 285, 532
Compton Gamma Ray Observatory 576
compulsory third-party liability insurance 829
computational fluid dynamic calculations 59
computer programs for damage prediction 111
Computer systems 473
concept phase 121
concurrent engineering 24
condensate 423
condensate water separator assembly 423
condensing heat exchanger 422
conditional sales contract 830
conducted emission 689
conducted and radiated emission 689
conducted and radiated susceptibility 689
configuration item data list (CIDL) 761
Configuration management (CM) 790
configuration tables 298
conical nozzle 314
conical scan feed 489
connecting roads 188
constant acceleration 446
constant drag 77
Constellation Program 506
Constellation Simulator (CSIM) 564
constellations 697
constraint satisfaction problem theory 481
Constraints on the mission design 656
consultants 751, 765
contact periods 468, 471
contact resistance 274
contact time 529
Contamination 50
contingency 298, 470, 759
contingency situations 298
continuous data collection 616
continuum flow regime 94
contract change note (CN) 767
contracts 823
Contractual Mitigation 759
control 295, 751
control accelerometers 673
control accounts 805
Control and monitoring unit 726
control and surveillance 361
control audit 774
control center 20
control center operations 468
control loop 340, 484
control loop time (latency) 22
control momentum gyro (CMG) 353
Control Rooms 473
Control segment 652
controllable antenna 650
controlled 589
Controlling 590

Convection 594
convective heat transfer 268
conventional temperature range 268
conversion of the primary energy 237
cooling 511
Cooling methods 321
cooling system 585
Cooling water 423
coordinate 813, 815
Coordinated Universal Time (UTC) 569
coordination of all needed subsystems 187
corona 36
coronal mass ejections 38, 49
correction maneuvers 77, 473
cosmic radiation 520
cost baseline 802
cosmic X-rays 575
Cost budgeting 802
Cost Breakdown Structure 799
cost considerations 670
cost control 797
cost drivers 798
cost elements 797
cost engineering 794
cost estimates 797
cost estimation methods and models 800
cost estimation relationships 801
cost for flight components 244
cost management 793
cost management process 795, 796
cost profile 802
cost reimbursement contract 803
cost and schedule 776
cost as an independent design variable 793
countdown 187
coupled, synergistic, biological systems 442
coupling devices (transducers) 689
course corrections 302
covariance analysis 463
covariance matrix 72
crack propagation 209
Craig–Bampton method 219
crater depth 108
credit agreement 830
crew qualification and responsibilities matrix 403
crew shift 450
criteria 772
critical 435
Critical design review (CDR) 122, 659, 703, 749
critical particle diameter 109
critical path 754
Criticality of Risk 759
cross-waiver of liability 827
cryogenic range 268
cryogenic stages 152
cryptographic components 559
CubeSat specification 702
current orbital position 483
current–voltage characteristic 250
curricula in space flight technology 3
customer product management (CPM) 752
customer questions 776
customer requirements 781
customer satisfaction 769
cycle life 258
cycloid 446
cylindrical shadow model 60
CZ-4B 126

D

daily observation duration 529
damage or ballistic limits equations 108
damage to satellites 762
damaging mechanisms 317
Darkening 47
Data access 471
data access in near real time 532
Data Libraries 534
data delivery interface document 472
data disposition system 471
data dissemination service 718
data distributor 820
data flow tests 465, 505
data management segments 361
data management system 512
data management system, attitude and orbit control system 467
Data Processing 19
data rate 590
data rates and volumes 532
data recall (replay) 478
data reduction 363, 373
data, video and audio connections 511
data, voice and video connections 516
database 480, 797
day–night variation 60
DC power distribution 723

Debris Mitigation Measures 101
decomposition chamber 317
decontamination 442
Decryption unit 726
deep dielectric charging 49
deep discharge 260
definition of the thermal system 278
Degradation 46, 48, 211, 217
degradation factors 250
Delays 484
deliverables 800
delivery in orbit 826
Delta IV-M 127
delta qualification 734
demodulated 485, 492
dendrites 600
density of the atmosphere 521
Deorbiting 638
departure 452
dependability 776
Deployable flexible or rollout arrays 255
Deployable rigid arrays 255
Deployable structures 214
Deployment Tests of Antennas and Booms 694
Deployment Tests of Antennas, Booms and Solar Arrays 695
depressurization maneuver 142
derivatives of the measurement model 70
Description of Consequence 759
description of the orbital path 61
design decisions 731
design drivers 203
design limit load 204
design of a rocket stage 149
Design of the Flight Control System 153
design of the transmission link 494
design phase 122
design philosophy 436, 660
design process 153
design to cost 190
design-to-budget 661
design-to-cost 661
design to minimum risk 437
design to objectives 660
design to science 660
design to value 190
Desorption 424
Detached spall 108
Detailed Definition Phase 749
detailed resolution 520
determination of longitude 553
determination of the attitude 473
deterministic attitude determination 340
development 122, 745
development and acceptance tests 668
development configuration baseline 659
Development Logic 166
development programs 741
development tests 198, 282, 665
deviation 569
demonstration or test flight 457
diagonal elements 72
diaphragm tanks 310
diesel-powered generator 477
Diet 613
differential charging 46
differential correction 574
diffusion node 279
digital links 484
digital systems 476
dimensioning load cases 204
dipole moment 339
direct access stations 533
direct energy conversion 241
direct energy transfer (DET) 245
direct interface force method (DIFM) 692
direct simulation Monte Carlo (DSMC) 94
direct television broadcasting 812
directional measurements 67
directional solidification 599
directness 744
discharge of static electricity (ESD) 689
discrete mass and surface elements 276
dispenser 730
display system 478
disposal 432
Disposal Phase 750
distance measurement 67
distribution of the maximum general proton flow 43
distributed single-point grounding 266
Disturbance Torques 338
disturbances 302
disturbing accelerations 692
Disturbing Signals and Their Propagation 689
docking 638
docking maneuver 444
docking mechanism 452

documentation 776
dog house effect 178
Doppler frequency 486
Doppler measurement system 492
Doppler shift 67
double wall 107
downconverter 491
Downlink 488
draconitic period 77
drag 77
drag coefficient 59
drift rate 79
drift velocity 447
drinking water 429
dry lubricants 232
dry mass 138
dual-gimbal CMGs 353
Dual-Mode 323
Dual-mode apogee engines 324
dual-use technology 816
ductile and brittle materials 108
dusk–dawn orbits 530
dynamic behavior 195
dynamic effects 593
dynamic interaction 626, 628
dynamic loading 216
dynamic loads 204
dynamic pressure 34
dynamo layer 40

E

EAC 797
earned value analysis 805
earned value management 804
Earth-Centered Orbits 55
Earth exploration satellites 485
Earth-fixed geocentric reference 335
Earth observation capacity 532
Earth observation and climate research 523
Earth observation satellites 17
Earth Resources Technology Satellite 524
Earth surveying systems 526
Earth swing-bys 586
Earth's Atmosphere 39, 82
Earth's gravitational potential 44
Earth's magnetic field 525
Earth's oblateness 345
easy handling 316
eccentric anomaly 54
eccentricity 53
eccentricity of the chaser orbit 446
eccentricity vector 79
eclipse 244
ECLSS 511
Economic Conditions 800
ECSS 761
edge nodes 276
Effect of Risk 759
effective emissivity 284, 285, 291
effective radiation background temperature 273
effective thermal conductivity 284
effective-to-isotropic radiated power 496
effectiveness 769
efficiency 176, 266
eigenfrequencies 203, 220
eigenmodes 220
eigenradiation 37
Einstein's theory of relativity 569
electric current 274
electric heater system 295
electric motor 226
electrical energy 18
Electrical Ground Support Equipment (EGSE) 664, 732
electrical mating condition 734
electrical power system 236
electrical resistance 292
electrolysis 420, 424
Electromagnetic propulsion 325
electromagnetic compatibility (EMC) 328, 689, 733
electromagnetic compatibility of radio systems 500
electromagnetic fields 688
electromagnetic field (radiated emission) 689
electromagnetic radiation 522
electromagnetic waves 701
electronic circuitry 319
Electrostatic propulsion 325
Electrostatic charging 46, 49
electrostatic thrusters 325
Electrothermal propulsion 318, 325
electrothermal hydrazine thrusters 318
elements of the space mission 647
elevation angle 487
embargo 816
Embrittlement 47
EMC (Electromagnetic Compatibility) 357, 736
emergency 470

emergency position-indicating radio beacon 730
EMI (Electromagnetic Interference) 357
emissivity in the infrared band 290
encryption 532, 726
end item data package (EIDP) 755
end of the operations phase 122
end to end 734
end-to-end operational scenarios 465
energy balance 269
energy flux density 37
energy or heat balance 273
Energy storage 238
energy supply 148, 292
engine housing 137
engine qualification 198
Engine Test Rig 694
engine thrust frame 146
engineering model 715, 731, 785
engineering model (EM) 703
Engineering Qualification Model 666
engineering and qualification model (EQM) 666, 703
engineering support team 511
engines that can be throttled 589
engines with turbopumps 174
environment 520
environmental conditions 668, 704
environmental investigations 529
environmental mapping 524
environmental mapping programs 525
environmental simulation tests 668
environmentally friendly 306
epoch by epoch 73
EQM 666
equation of motion 678
equation of motion for the satellites 560
equivalent dose 48
equivalent isotropic radiated power 490
equivalent noise temperature 498
ergonomic design 474
ERIS uplink station 568
erosion 47
error correction 502
ESA 832
ESA implementation phases 747
ESA MedOps 514
ESA standard mode 724
escape velocity 587
estimate to complete 806
ETC 797
ethylene 440
EU 832
EU terrorism list 816
Euler angles 336
Euler equations 98
euphoric phase 406
EURECA 13
European Centre for Medium-Range Weather Forecasts 524
European Cooperation for Space Standardization 747
European Infrared Science Observatory 576
European interconnection ground subnet 514
European ISS ground infrastructure 513
European Joint Operations Panel 510
European modular cultivation system (EMCS) 442
European Space Agency 7
European Space Policy 833
European space programs 8
European Terrestrial Reference System 1989 570
evaluation 776
exercise of remedies 830
exhaust nozzle 318
exhaust velocity 119
exosphere 40
expander engine 158
expansion of the Universe 577
expected value 72
expeditions 402
expendable launch vehicles 122
Experimental demonstration 192
experiments 506
exploration of the Solar System 575, 697
Explorer 1 519
Explorer 11 576
exponential function 259
Export Control 816
External loads 178
external pressurization 157
external service provider 564
external system 493
External torques 338
extra EMC 689
extraterrestrial material 521
extravehicular mobility unit 431
extreme requirements 116

F

face sheets 207
facilities and equipment 769
failed heater 300
fail-safe mode 725
failure analysis 470
failure conditions 236
Failure management 791
failure of a launcher 194
Failure prevention instead of failure correction 770
Failure tree analyses 194
fairing 146
Falcon 1 124
Far-field 383
far-range approach 443
Faraday cages 736
fatigue 178
fatigue failures 191
fault detection, isolation and recovery 357
Faulty thermistors 300
FCL 267
feasibility 24, 749
Feasibility Study 751
Federal Office of Economic and Export Control 816
feed losses 497
FGB1 423
fiberglass composite 215
fiber–metal laminates 206
fictitious force of inertia 592
field emission thruster 326
figure of merit 287, 498
filament winding technique 206, 208
filler materials 275
filters 152, 330
final data 561
final approach 445
final design 278
financial budget 24
fine structures of the signals 719
Fine Sun sensors 344
finely woven mesh of subelements 281
finite difference and finite element methods 278
finite difference method 219
finite element method 217, 281
finite element model 203
fire extinguishing chemical 430
fire prevention 436
Firewall and proxy systems 477
firing test 198
Firing Test for Solid Propellant Boosters 693
firm fixed price 765
fixed alkaline electrolysis 424
fixed price 805
fixed price contract 803
Fixed satellite communication 552
Flagging 478
flammable segment 430
flattening, Earth 57, 58
flight conditions 192
flight control 146
flight dynamics 472
flight dynamics calculations 473
Flight Dynamics Facility (FDF) 564
flight engineer 404
flight model 223, 731, 758, 785
flight model (FM) 703
flight of the first stage 140
flight operations 20
flight operations plan 460
flight qualification review 659
Flight readiness review (FRR) 659, 750
flight software 733
flight tasks 403
flight-standard components 732
floating orbit 530
flow control unit 328
fluid displacement 611
fluid heat carrier 286
flush-mounted inserts 213
flushing toilets 434
flux concentration factor 106
fly-by missions 521
flywheels 230, 241, 257
Focal length 342
focusing factor 104
foil radiators 294
food storage 432
Forces of Solar Radiation 560
formation flights 531
formation flying 649
forward error correction 729
four main engines of 490 N 454
four satellites in space 732
frame of reference 54
Framework Agreement 833
Free Space Attenuation 499, 500
free surface 594
Freon 432

frequency allocations 495
Frequency and Time Reference 492
Frequency Assignment 813
frequency band 495
frequency ranges 494, 531, 671
frequency spectrum 494
Frequency Utilization 543
Fresnel region 383
Friction 231
frozen orbit 530
FS 666
fuel 303
fuel cells 242, 257
full flight standards 732
full operational capability 737
functional analysis 777
functional baseline configuration 658
Functional requirements 654
functional requirements specification 190
Functional tests 668, 693
functioning of a booster 693
function tree 798
fundamentals of human space flight 407
Further models 716
fuse blowing 247

G

gain factor 497, 498
gain figure 498
Galactic cosmic rays 49
Galileo 520, 557
Galileo control centers (GCC) 723
Galileo navigation system 28
Galileo reference system time 568
Galileo satellite navigation system 8
Galileo sensor stations (GSS) 562
Gamma Ray Astronomy 576
gas chromatograph separator 426
gas generators 160
gas radiation 88
gaseous pollutants 425
Gaussian variational equations 61
GCS 556, 562
general introduction 407
generic phase model 781
GEO (Geostationary Orbit) satellites 556
geocentric and interplanetary trajectory 463
Geodesy and Cartography 554
geographical distribution 800
geoinformation services 536
geomagnetic and solar incidents 47
Geometric resolution 526, 528
georeturn rule 117
geostationary communications satellite 534
geostationary Earth orbit 540
geostationary transfer orbit 648
geostationary weather satellites 524
geosynchronous orbital radius 78
German Space Agency 824
German Aerospace Center 821
German Aviation Authority 822
GETEX 628
geyser effect 141
GF value 281
GL value 280
glass fibers 206
global ground station networks 21
global measurement procedures 570
Global Monitoring of Environment and Security program 536
global positioning system 28
GLONASS 557
GMES program 536
GMS 556, 562
GNSS receivers 573
Google Earth 520
gossamer structures 216
Governmental and security-sensitive user 719
GPS 557
GPS time 569
GPS/Galileo receivers 573
GR value 281
Grants for Space Missions 821
graphical editor 481
graveyard orbits 102, 470, 649
gravitational losses 120
gravitational acceleration 53
gravitational field coefficients 58
gravitational field models 57
gravitational perturbations 57, 79
gravitational potential 559
Gravity 43, 522, 525
Gravity field 525, 588
gravity-assist 588
Green Propellants 306
Ground-based testing facilities 83
ground communication infrastructure 506

ground control points 526
ground operations 459
ground resolution 524
ground segment 16, 20, 121, 195, 652
Ground Station 464, 468, 484
ground station networks 533
ground support equipment (GSE) 762
ground system 658
Ground System Validation and Testing 464
ground vibration test 678
Guidance, navigation and control 121
gyroscopic effect 348

H

H-IIA 127
H-KC 12 GA 317
half-sine shocks 676
Hall Effect Thruster 326
HAN (Hydroxylammonium Nitrate) 307
hard radiation 343
Hardening 210
Hardware in the Loop Test 359
hardware matrix 667
hardware-in-the-loop simulation 716
health of the European astronauts 517
health risk 442
heat conduction and diffusion 87
heat exchanger module 422
Heat Flux Balance 87
heat flux density 274
heat loads 142, 423
heat pipes 286, 295, 688
heat pipe architecture 737
heat pipe liquid 287
Heat radiation 294
Heat transport 274
heat transport processes 274
heater layout 294
heating elements 283, 589
Heavy-lift launch vehicles 123
hemispheric resonator 349
Hermann Ganswindt 4
Hermann Oberth 5
Herschel–Planck 577
Hertzian stress 232
heterosphere 40
Hi-Eta-Si cell 251
high concentration 316
high flow rates 312
high geometric resolution radar systems 527
high reliability 514
high thrust domain 160
High Vacuum 34, 49, 50, 211
high-efficiency particle filter 425
High-Energy Particles 34, 39, 211
High-energy radiation 211, 588
High-frequency disturbances 689
high-gain antennas 590
high-power amplifier 490
high-pressure helium tank 303
high-pressure xenon bulbs 688
high-temperature catalytic converter 425
high-temperature systems 268
High-voltage technologies 236
highest precision 573
Hohmann maneuvers 445
Hohmann orbit 648
Hohmann Transfer 75, 448, 587
hold point 447
homosphere 40
honeycomb core 207
horizon profile 487
hot backup 478
hot timeline 294
hotspots 531
house keeping data 484, 485
Hubble Space Telescope 577
Human missions 117
human mission to Mars 607
human Moon landing 9
human safety 762
human space flight missions 505
human space operations 506
humidity sources 422
Hybrid Bus 249
hybrid materials 206
hybrid propulsion systems 154
hydrazine 303
hydrazine arcjets 318
hydrazine gas generator 318
hydrogen masers 557, 728
Hydrogen peroxide 303, 305
hypersonic flows 82
hypersonic testing 96
hyperspectral resolution 520, 531
hypervelocity impacts 108

I
identification 776
ignition process 140
ILS concept 761
images of celestial bodies 473
imaging radar sensors 528
immunity level 689
impact risk 104
impact risk due to space debris 106
impact velocities 111
impacts of particles of a certain size 106
Implementing Attitude Control 695
improving orbital parameters 66
impulse bit 305, 314
impulse-type maneuver 77
in-house rules 761
in-orbit testing 21
in-situ investigations 521
inadequate storability 305
inclination 54
inclination vector 79
Increased charge current 260
Increased discharge current 260
increments 510
independent cost estimates 797
independent power connections 512
Independently Performed Functional Tests 693
indications of life 582
indirect interface force method (IIFM) 692
induced atmosphere 35
industrial PC 491
Inertial geocentric reference system 335
inertial measurements 341
inertial velocity 56
infrared 36, 521
Infrared sensors 531
infrared wavelength region 270
infrastructure 488
infrastructure of a launch site 184
ingestion processing sequence 535
initial velocity 448
injecting a liquid 149
injection head 317
Injection orbit 648
injector 320
INMARSAT 539
Inserts 212
inspection methods 192
instant position measurements 69
Institute for Cosmos Research 6
Instructional system development 410
insurance 828
Integral 576
integrated logistic support (ILS) 761
Integrated Logistics Support Tools (ILS Tools) 564
Integrated Spacecraft Tests (IST) 736
Integrity 762
integrity data 726
intellectual property rights 824
INTELSAT 539
interdiffusion 601
Interface to the Customer 757
interface to the space segment 565
interfaces 33, 196, 506, 510, 532
Interfaces to Internal Management 757
Interfaces to the Project Team 757
Interference 500
interference limited 503
interferometric methods 530
Intermediate-Level Runs 674
internal pressurization 157
Internal auditing 769
internal LAN 564
internal magnetic moment 45
Internal torques 338
international astronaut classes 408
International Celestial Reference System 56
international cooperation 506
international distress calling frequencies 726
international register 810
International Space Station 505
International Terrestrial Reference System 56
interoperability 568, 761
interplanetary probes 301
interplanetary space flight 29
interstage structures 145
intertank structures 145
intra EMC 689
invariable orbital plane 52
ion exchange beds 427
ionizable gases 303
ionization 42, 325
ionosphere 574
ionospheric effect 500
IRAS 576
iridium 316
irregular form of the Earth 79
irregular radiation distribution 345

ISO 761
isoflux characteristic 720
isogrid structures 209
ISS crew surgeons 405
ISS partnership 402
IT security 477

J
Jacobi matrix 71
JEM 423
Jupiter 584
jurisdiction and control 810

K
Kalman Filter Orbit Determination 73
Kalman filtering 66
Kalman gain 74
Kaufman thruster 326
KC 12 GA 317
keep out sphere 450
Kepler Orbits 52
Keplerian elements 55, 559
Kevlar 206, 215
Key Inspection Point (KIP) 754
Keyhole 528
kick stages 136
kill or deactivate microbiological loads 442
Kirchhoff radiation law 271
klystron 490
Knudsen number 85
Konstantin E. Tsiolkovsky 4

L
labeling structural parts 214
laboratory 404
Lagrange points 44
land surface 524
Lander Missions 632
landers 582
landing 589
landing people 521
Landing Systems 589
Landsat 1 524
large data archives 532
large parabolic antennas 590
laser ranging 66
laser tracking 525
latch current limiter 295
laughing gas 307
launch and early operation phase 21
Launch and Early Orbit Phase 467
Launch Cost 542
launch element 648
launch locks 216
launch mass at the beginning of a mission 721
launch operator 118
launch readiness review 659
Launch Rehearsal 694
launch segment 652, 653
launch service agreement 826
launch systems 116, 121, 721
launch vehicle configuration 193
Launch vehicle 115, 196
launch windows 587, 588
launched horizontally 123
launched vertically 122
launcher analysis 149
launcher system 6
launching state 809, 810
laundry 422
laws of celestial mechanics 52
laws of friction 231
LCL 267
Leakage 152
leasing contract 830
Least Squares Adjustment Orbit Determination 71
least squares estimation method 340
legacy mode 730
LEO (Low Earth Orbit) constellations 556
lessons learned 768
level of autonomy 460
Liability 831
Liability Convention 810
liability for damages 810
license to distribute remote sensing data 821
life cycle 745
life cycle cost 794
life cycle of a space project 747
life cycle of space products 747
life support 29
life support systems 610
lifetime 470
lifetime of a mission 25
lift/drag ratio 91
light arc 319
lightning 689
Lightweight Directory Access Protocol 477

Lightweight structures 133
line organization 758
line scanners 530
linear differential equations 446
linear energy transfer 49
linear power amplifier 726
link 494
Link Margin 504
Linking 760
liquid carry-over sensor 423
liquid lubricants 232
liquid propellants 136, 154
LISA 523
Lisbon Treaty 833
lithium hydroxide 431
lithium hydroxide bed 425
lithium-ion accumulators 701
lithium-ion cells 257
load-bearing structure 722
load cases for qualification tests 193
load profile 684
load requirements 153
LOC 372
local area networks 493
local fields 521
local user terminals 730
Location of Stages 136
logistic support analysis (LSA) 761
long direct injection ascent phase 723
long-term orbiting systems 29
loop filter bandwidth 491
loop heat pipe 289
loss figure 498
louvers 283, 290
low Earth orbits (LEO) 698
low thrust 303
low-level sine test 222
low-level sine vibration tests 220
low-noise amplifiers 485
low-pressure transducer 313
low-temperature (LT) bus 431
lower stages 136
lowest possible integrated level 193
lumped parameter method 279
Luna 9 579
Luna 16, 20 and 24 580
Luna 17 and 21 580
lunar and Mars exploration 115
Lunik 2 579
Lunik 3 579
Lunokhod 1 and 2 580

M

Mach number 93
magnetic bottle 42
magnetic cleanliness program 690
magnetic dipole 41, 356
magnetic effects 690
magnetic field 346
magnetic field simulation assembly (MFSA) 690
Magnetic measurements 670
magnetic models 45
magnetic substorms 38
magnetic system test 692
Magnetism 43, 45
magnetorquers 230
magnetosphere 40
main and upper stage 155
main bus 245
main contractual obligations 823
main control room 467
main engine 445
main lobe 499
main propulsion 303
main structural elements 153
maintainability concepts 778
Maintaining the quality 769
maintenance 511
management decision 741
management of margins and risks 191
Managing/displaying 478
Mandatory Inspection Point (MIP) 755
maneuver 586
maneuver planning 80
manned Apollo missions 580
manual commanding 471
manually controlled 444
manually controlled docking 443
manufacturing 122, 674
manufacturing technology 173
mapping the Martian surface 582
margin of safety 219, 220
margins 118
Mariner 2 581
Mars 581
Mars swing-by 586
Martian soil 582
mass 327

mass budget 150
mass distribution 204
mass flow 162
mass loss 49
mass matrix 681
mass model 714
mass properties 671, 681
mass spectrometer 421, 426
mass-efficient shield arrangement 110
master science plan 470
material compatibility 316
materials 283, 307
materials science 591
mathematical model 691
mating of two Earth-orbiting systems in space 443
matrix organization 758
maximum power point 250
maximum power point tracking (MPPT) 246
maximum velocity increments 304
mean drift 79
mean geographic longitude 79
mean time for repair (MTFR) 761
measure the change in distance 67
measurement update 73
measurements through clouds 531
measuring conductivity 430
Measuring the tank temperature 313
Mechanical and structural approval 733
mechanical ground support equipment 205, 214
Mechanical Ground Support Equipment (MGSE) 664
mechanical shock 260
mechanical stresses 671
mechanical structure 18
mechanical systems 191
mechanisms 203, 223, 230
Medical care 432
medication 441
medium Earth orbit (MED) 718
Medium launchers 123
medium to very high geometric resolution 529
medium-range rocket 5
MEO (Medium Earth Orbit) satellites 556
mercury 306
mesosphere 39
metal matrix composites 206
metallic diaphragms 311
metallic screens 288
metallic structures 209
Meteoroids 100
meteorology climatology 529
methane pyrolysis 435
method of concentrated parameters 279
method of least squares 66
methodological competence 742
methodological repertoire 610
microbial growth 422
microgravity 522, 591
microgravity, aerodynamic drag 34
micrometeorites 33
Micrometeoroids 51
microporous molecular structure 285
microsatellites 706
Microvibration 692
microvibration budgets 692
microvibration measurements 692
microwave background 521
microwave range 521
microwave sensors 530
middleware 373
MIL 1553B data bus 724
Milestone Plan 754
milestones 746
military and civilian missions 117
military considerations 519
military end use 816
military reconnaissance mission 528
miniaturization 308, 697, 698
miniaturization of instruments 698
minimal interference 728
minimize risk 471
miscibility gaps 599
mission analysis 118
Mission Analysis and Systems Studies 463
mission architecture 653
mission concept 24, 653
mission control center 651
Mission control rooms 474
mission control team 470
Mission definition review (MDR) 749
mission evolution 463
mission implementation plan 464
Mission Objectives 654
Mission operations 651
Mission planning 473, 651
Mission profiles 463
Mission receiver 726

mission requirements 654
mission specialists 403
Mission Statement 653
mission-unique tests 465
Mixed burners 161
mixture ratio 156, 324
MMH 306
M&C system 505
mobile communications 520
Mobile telecommunications 552
Mobility aids 432
modal analysis 679
modal analysis of vibration test data 680
modal mass 220
modal models 219
modal parameters 678
Modal Survey 678
modal test 679
model function 70
model of the measurement process 69
model philosophy 666, 669, 745, 772
model updating 220
model verification 219
moderate-temperature (MT) bus 431
modulation 551
modulation index 386, 389
module 730
mold 422
moments of inertia 204, 682
momentum 588
momentum management 356
momentum part 155
momentum wheels 230, 353, 722
MON 1 306
MON 3 306
monitor 751
Monitoring and control 534
monitoring and control system 504, 508
monitoring and safety unit 454
Monitoring the Oceans 527
Monitoring the temperature 313
monocoque constructions 145
Monopropellant 303, 725
monopropellant or bipropellant systems 307
monopulse tracking system 489
monotony 616
Monte Carlo simulations 359
Moon Treaty 812
MoS 219
most cost-effective 302
motion of a spacecraft 33
motion of artificial Earth-orbiting satellites 52
motion simulator 686, 688
motor case 162
mounting of the HPA onto the feed system 490
moving objects 528
MPPT regulation 255
multibody systems 44
multidipole model 692
multijunction GaAs solar cells 701
multilayer insulation 107, 283, 588
multimission environment 25
multimission ground segments 533
multipath effects 351
multipath propagation 574
multiple frequencies 570
multiple polarization capability 531
multiple reignition capability 141
multiple stage 122
multisatellite launches 730
Muscles 612
N_2O_4 306
nadir pointing 725
nanosatellites 697
NASTRAN 218
national registry 809
national security or foreign policy interests 817
national space flight program 6
national space legislation 818
natural atmosphere 34
nature of the objects that can be observed in the nocturnal sky 575
Navier–Stokes equation 98
navigation 638
Navigation and Flight Dynamics 472
navigation antenna 726
navigation data 570
navigation messages 561, 573, 726
navigation satellites 17
navigation signal generator 734
navigation signal generator unit 558
navigation signals 718
navigation system 68
navigation terminal 556
near real time 535
NEAs 225
negative pressure relief valves 420
Neptune 584

network model 278
Networking technologies 474
Neutral burners 161
neutral gas atmosphere 47
Newton method 54, 93
nickel–cadmium 257
nickel–hydrogen 257
nitrogen 305
NOAA series 523
node model 294
node temperatures 276
nodes 217, 276
Nodes 2 and 3 423
Noise and odor control 433
noise contribution of the total link 504
Noise fields 674
noise temperature 498
noise temperature of the system 498
NOM1 table 298
NOM2 table 298
NOM3 table 298
Nomex 206, 207
nominal operating pressure 308
nominal operation limit 294
non-conformance 791
non-disclosure 824
non-military space flight 1
noncatalytic surface 87
nonconformance reports (NCRs) 767
nonexplosive actuators 225
nonlinear properties 352
nonoperational limits 294
nonregenerative systems 434
nonspinning satellites or probes 310
nonstructural mass 219
NORAD orbital elements 62
normal equations 72
nose of the rocket (fairing) 151
notching 223, 674
nozzle 162
nozzle throat 321
Nuclear fusion processes 35
nuclear power sources 812
Nuclear power supply systems 241
nucleus and environment of an active comet 586
number of engines 142
number of launches 116
number of nodes 276
number of stages 119
numeric simulations 108
numerical flow field analyses 96
Numerical Orbital Predictions 64
nutation 56
nutrition source 440

O

obligations contractual 831
obligations for 809
observation areas of interest 529
observation of the Earth's surface 529
observation profile 529
obsolescenses 198
occurrence probability of failure 776
ocean topography 528
off-line approach 460
office communication system 493
omnidirectional antenna characteristic 724
on-board antennas 736
On-board clock accuracy 728
on-board control 443
on-board computer 361
on-board orbital control 78
On-board Processing and Switching 550
on-orbit short-term plan 510
on-orbit summary 404, 510
on-stage thermal cycling (OSTC) test 684
one-failure tolerant 454
online access services 535
Online access via the Internet 534
open service (OS) 571
operating parameters 693
operating point 497
operating pressure 307
operating systems 475
operating temperature range 286
operating temperatures 260
operating voltage range 259
operation effort 25
Operation Preparation Facility (OPF) 564
Operation (LEOP (Launch and -Early -Operation Phase) and Operation Phase) 750
operational experience 133
operational limits 294
operational mission requirements 654
operational range 322
Operational readiness review (ORR) 659, 750
operational requirements 655
operational software 467

operations 506
Operations and Service Phase 752
operations coordinator 513
operations phase 122
operator of an advanced remote sensing satellite system 820
OPS-LAN 493
OpsWeb 483
Optical alignment 736
Optical cameras 530
optical components 688
Optical Ground Support Equipment (OGSE) 664
optical range 527
Optical sensing systems 531
optical sensors 530
optical solar reflectors 290
optics 577
Optimized customer 770
optimum operating point 248
optimum orbit 23
orbit of a space mission 648
orbit adjustment 588
orbit and attitude control 303
orbit around Lagrange point 462
Orbit Control 301
orbit determination 473
orbit displays 483
orbit insertion maneuvers 591
orbit maneuvering system 445
Orbit reference system with its origin in the spacecraft's center of mass 335
orbital control maneuvers 79
orbital elements 55, 68
orbital maneuvers 75
Orbital parameters 244
orbital periods 53
Orbital Perturbations 57
orbital positions 724
orbital reentry 85
orbital system 506
orbiters 582
Order management 534
ordnance devices 223
organic substances 426
organizational structure 468
organizations 3
organization's processes 769
organizing quality management and product assurance 775
orientation 19
orthogonal (Cartesian) coordinate 720
Oscillating rotation rate sensors 349
OSO-3 575
Outer Space Treaty 809
outgassing 49, 211, 683
output backoff (BO) 497
output multiplexer 726
overall schedule 759
overall sound pressure level 675
overall volume 721
Overcharge 260
overdetermined system of equations 73
Overlapping 285
overpressure valves 419
overworking 405
oxidizer 303
oxidizer-compatible materials 310
oxygen cartridges 434
oxygen concentration 421
oxygen cycle 424
oxygen masks 431
oxygen sensor 421
ozone hole 520

P

P-POD (Poly Picosatellite Orbital Deployer) 698
packet error rate 494
panchromatic sensor 530
parameter studies 273
parameter variation 278
Parametric cost models 801
Parking orbit 648
particle filter 422, 426
particle fluxes 521, 522
particle shape 111
particle size 111
passivation 102, 140
passive intermodulation (PIM) 488, 736
passive laser gyro 349
passive laser ranging reflector 725
passive microwave radiometers 532
passive protection element 266
passive system 555
payload architecture 726
payload control center 651
payload data 485
Payload Ground Segment 532
payload instruments 468

payload mass 154
payload module 730
Payload Operations and Integration Center 513
Payload preparation 185
Payload science operations 460
payloads 17, 121, 292, 468, 506, 584, 591, 649, 719
payment 831
PCM mode 492
Peak data rates 532
Pegasus-XL 124
percentage of work complete 804
perforations 285
performance parameters 733
performance tests 693
Performance Tests of the Electronics 694, 695
perihelion 587
permanent stand-by operation of the high-precision clocks 732
permanently occupied station 402
permeability 310
Personal hygiene 432
Personnel Measures 762
perturbation theory 61
perturbations 79
Phase A study 462
Phase Resonance Technique 678
Phase Separation Technique 679
phased array antenna 549
phasing 443, 444, 445
Philae 586
Photoelectric Effect 46
photogrammetry 526
photosynthesis 440
photovoltaic 237
photovoltaic effect 46, 249
physical availability 494
physical models 108, 358
physical phenomena 27
physical presence 521
physical progress measurement 804
physical separation 185
physical vapor deposition 233
physics of black holes 577
Phytotrons 439
picosatellites 696
piezoceramic materials 228
piggyback launch strategy 706
pinpullers 224
Pioneer 4 579
pitch angle 336
pitch axis 346
Pixel resolution 342
Planck radiation law 36, 271
planetary gear 227
Planning 510
planning of the product assurance task 787
planning tasks 760
plasma volume 619
plastic deformation 686
plastic foil 284
platform 719, 730
playback mode 462
POGO Oscillations 179
Point-to-point architecture 650
pointing of large antennas 499
pointing accuracy 313
polar data-dump stations 533
political stability 184
pollutants and metabolites 440
pollution 151
position, velocity and time 68
positioning and attitude control 147
positions in the control room 507
positive pressure relief valves 420
post-pass report 505
potting 212
Power 582, 584
Power conditioning 238
Power conversion 238
Power distribution 238
power generation 455
power lockup 246
power management and distribution (PMA) 245
power requirement of the turbine 173
power requirements 244
power spectral density 223, 673
power supply 25, 511, 734
power system 333
Prandtl number 84
precession torque 348
precise mode 332
precision of the attitude control 526
predicted temperature variation 294
prediction of satellite orbits 52
predictor–corrector method 65
Preliminary definition review (PDR) 749
Preliminary Definition Phase 749
preliminary design phase 122

preliminary design review 122, 659
Preliminary requirements review (PRR) 658, 749
prepregs 206
Presence in Space 521
preshipment review 659
pressure control assembly 329
pressure-compounded 177
pressure equalization valve 420
Pressure fed systems 143
pressure-fed 174
pressure part 155
pressure-regulated 308, 323
pressure transducers 329
pressurization gas 145
pressurization system 150
pressurized gases 143
preventing product failures 768
primary batteries 237
primary cells 257
primary electric arc discharges 256
primary ground station 650
primary power 721
primary structure 203, 584
principal axes 335, 681
principles of system design 658
Private and commercial users 719
PRN code 570
Probability of Risk 759
probability of survival 111
probe and drogue 452
procedure to authorize 817
procedures of a mission 506
process noise 74
process qualifications 769
process steps 760
processing 478, 485, 534
processing and archiving facilities 652
procurement program 745
produce food 439
product and project audits 776
product assurance 755, 761, 787
product assurance and safety (PA&S) 762
product assurance plans 776, 781
product assurance task flow 788
product assurance tasks 776
product life cycle 787
product tree 798
product-specific processes 772
product-specific processes needed for product generation 769
production configuration baseline 659
production of assets 741
Production Phase (MAIT Phase - Manufacturing, Assembly, Integration, Testing) 749
Production planning 774
production process 198
Production Process/Procurement Phase 752
production rate of ions 40
Program segment 652
program tracked 499
Progressive burners 161
Project Activities 788
project associated modeling (V-Model) 781
project breakdown structure 798
project close-out 807
project control 757, 796
Project Leadership 756
project requirements 484
Project Schedule 800
Project Structure (Work Flow Plan) 752
proof of concept 737
proof pressure and leakage tests 330
propellant classification 156
propellant combination 141, 156, 320
propellant feed system 174
propellant isolation assembly 329
propellant loading activities 331
propellant mass 138, 315
propellant orientation method 310
Propulsion 115, 301
Propulsion System 292, 722
Propulsion system performance 133
Protection of Spacecraft Against Impacting Particles 107
protective cover panel 216
protective shield 107
Proton 130, 721
proton exchange membrane 424
Protoplanetary disks 577
prototype flight model 715, 731, 785
prototypes 703
protruding inserts 213
PSD 675
pseudo-random noise (PRN) code 570
pseudo-range 573
pseudo-range measurements 555, 573
psychological acceptability 442

public regulated service (PRS) 572
public–private partnerships 1, 826
public tender 825
pulsed plasma thrusters 327
Pulse width modulated 265
pump-fed 174
pump fed systems 143
pump-fed engines 158
pump speed 175
purging 213
purity 316
pyrobolts 223
pyrogen tests 442
pyroknife 223
pyrolysis 425
Pyromechanical separation mechanisms 223
pyroshock spectrum 676
pyrotechnic systems 148
pyrotechnic valves 329
pyrotechnical excitation 676
Pyrotechnical shock simulation 677

Q

Q-level 222
QPSK modulation 389
qualification 97, 668
qualification and acceptance program 670
qualification commission 122
qualification level 222, 675
qualification of the propulsion systems 198
Qualification review (QR) 659, 703, 750
Qualification Runs 674
Qualification tests 192, 282, 665, 668
qualified and certified quality management system 793
quality assurance 761
quality assurance measures 769
quality control loops 769
quality controls 769
quality documentation 769
quality management for space products 768
quality management systems 747
quality of the distributed power 244
quality of transmission 503
quality plans and work instructions 769
quality processes and product processes 772
quality standards 769, 778, 779
Quartz SSMs 290
Quasi-harmonic loads 671
quasi-Keplerian model 560
quasi-static 204
quasi-static design load 204
Quasi-static loads 223, 671

R

R-bar 444
R-bar hopping 448
radar monitoring 520
radially oriented thrusters 302
radiate all dissipated heat 720
radiated susceptibility 689
radiation 268
radiation behavior 283
Radiation Belt (Van Allen Belt) 42
radiation belts 48
radiation characteristic 499, 736
radiation cooling 89
radiation dose 48
radiation energy 36
Radiation monitoring 432
radiation pressure coefficient 60
radiation pressure of the Sun 60
Radiation Protection 619
radiation temperature 37
radiative heat exchange 283
radiator surfaces 290
radiators 269, 589
Radio and television broadcasting 552
radio astronomy service 729
radio frequency contacts 460, 467
radio link test model (RTM) 703
radio services 495
radio transmitters 68
radioactive heater units 588
radioactive radiation 48
radiofrequency carrier 484
radiofrequency compatibility 464
Radiofrequency Ion Engine 325
Radiofrequency Ion Thruster 242
radioisotope thermoelectric generators 241
radiometric data 473
radiometric methods 66
radiothermal generators 582, 588
Rain Attenuation 500
random spectra 204
random dynamic loads 213
Random errors 342
Random excitation 672

random load 223, 671
range safety 147
Ranger 580
ranging 485
ranging measurement 492
ranging measurement system 492
Rankine process 240
rapid data 561
RapidEye system 524
reaction control subsystem 467
reaction degree 177
reaction wheels 352, 353
real-time navigation 65
real-time navigation system 68
real-time processing 73, 562
reasons for microvibration 692
receiver unit 650
recirculating fan 421
recovered mass loss 211
Recovery Indicator 759
reduced data rates 533
reduced stiffness matrix 219
reduction of orbital energy 59
redundancy 295, 319, 362, 365
redundancy concepts 244
redundant systems 436, 440
Reed–Solomon (RS) convolution code 724
reentry corridor 93
reentry of a spacecraft 83
reentry path 90
reference frequency 493
reference spheroid 570
reference systems 334
Regeneration 424
regeneration rate 434
Registration Convention 809
Registration of Space Objects 809, 822
Regulated Bus 246
Regulation (EC)1334/2000 816
regulation options 266
reignitable stages 136
relative velocity 47
reliability 305, 818
remote programming 626
remote sensing 812
Remote terminal units 726
remove the before-flight pin 702
rendezvous 586
rental or leasing contract 828
Reorbiting 638
repairing faulty parts 511
repeat cycle 530
repeat ground track 26
repeat orbit 77
replacement system 436
repressurization 308
reprocessing of data sets 532
repulsion principle 115
Request for Information Phase (RFI) 751
Request for Proposal (RFP) Process 752
Rescue Agreement 811
research and development contract 825
residual atmosphere 338
residual force of gravity 593
Residual gases 285
resin transfer molding 210
resistance testing 736
resistance heater 318
resistojet 314
resolutions 812, 820
resonance behavior 203, 220
resonance characteristics 671
Resonance Search Run 674
resonance terms 62
resonances 179
resource rights 402
respiration index 434
responsible 809
Reusability 131
reusable launch system 133
reusable launch vehicles 122
reverberation chambers 675
reversal 311
review board 749, 766
review data package 754, 766
Review identified discrepancy (RID) 766
Review Planning 754
review process 754
reviews 746, 749
revisit time 530
Reynolds number 86
RF characteristics 736
RF shielding 732
RF transmission frequencies 726
RF Kurs 445
right ascension (or longitude) of the ascending node 54
rigid beam elements 218

ringing plate test setups 677
Risk Item Description 759
risk management 769, 800
risk matrix 759
Risk Owner 759
risk register 800
risk resulting from meteoroids 104
risks 829
Robert H. Goddard 5
rocket equation 155
Rockot 124
rod elements 217
Roll 336
roll axis 346
ROSAT 576
Rosetta 586
rotating, Earth-centered system 56
rotating mirror 345
rotation rates 341
rotational measurements 692
rotations in an inertial reference frame 347
rotor 176
rotor dynamics 175
roughness 275
routine operations 468
routine phase 406
RTGs 325
rubber diaphragms 310
rubidium atomic frequency standard 728
Rudolf Nebel 5
Runge–Kutta fourth-order method 65
rupture 178
Russian experiment operations 513
Russian On-Orbit Segment 437
Russian system 403
RVD camera 445

S

S-band 590
Sabatier reaction 425
safe mode 332
safety 818
safety analysis 762
safety document 778
safety equipment 331
safety factors 205
safety margins 192
Safety of Life (SoL) users 719
safety of life service (SoL) 571
safety precautions 505
sales contracts for products and services 828
Salyut 6/7 9
same sequence 670
sample return missions 51
sandwich analysis 207
sandwich core 207
sandwich structure 107
SAR antenna 730
SAR payload 736
SAR-Lupe 752
SATCOM Bw system 752
satellite and attitude control systems 155
satellite attitude 339
satellite bus 649, 719
satellite classification 706
satellite cluster 649
satellite communication and navigation 2
satellite communications 538
satellite constellation 649
satellite formation 649
satellite interfaces 733
Satellite Laser Ranging 67
satellite phased array antenna 552
satellite platform 17
satellite system 658
satellites 17, 649
Saturn 584
Scanning Earth Sensors 345
schedule 505
Schmidt number 84
science mode 332
science objectives 470
science operations planning 463
Scientific and technological support 651
scientific community 2
Scientific data 472
scientific or economic purposes 519
scientific problems 505
screen attenuation 689
search and rescue (SAR) 572
search and rescue payload 734
second 568
second surface mirrors 290, 588
secondary batteries 239, 257
Secondary particle radiation 49
secondary payload 698
secondary sustained arc 256
secular node drift 76

Security 762
security and foreign policy interests 820, 821
security requirements 474
segments of a GNSS 652
segments of an Earth observation mission 652
Segments of a Space Mission 652
self-decomposition 316
self-decomposition rate 316
self-diffusion 601
self-insurance 828
semi-major axis 447
semi-minor axis 447
Semiregulated Bus 248
sensitivity check 821
sensitivity matrix 71
sensor head 347
sensors 526
Sensory Systems 613
separation and departure 444
sequential lobing 489
series production 731
Series regulation 246
service level agreement (SLA) 752
service providers 751
service quality 503
service segment 652
several stages 136
shape memory alloys 228
shear forces 594
shelf life 258
Shell 405 catalyst 305
shell elements 217
shielding 104
shirtsleeve environment 512
shock generating units 676
shock layer in front of a reentry vehicle 86
shock loads 204, 223
shock response spectrum (SRS) 676
shock tunnels 96
shock waves 224
shocks, vibrations and acoustic noise 153
shunt diode 252
Shunt regulation 246
shutdown 264
sidereal time 55
signal availability 494
signal generators 555
Signal Latency 543
signal propagation time 589
Signal strength 342
signals in the time domain 726
silicon carbide compounds 206
simulation programs 135
simulations 457, 509
simultaneous lobing 489
sine load 222
single event phenomena 48
single or multistage pump 174
single point failure 226
single stage to orbit 122
single-component systems 160
single-gimbal CMGs 353
sinusoidal spectra 204
Sinusoidal excitation 672
six isolated thermal zones 722
Skeletal System 612
Skylab 47
slew maneuvers 333
Sloshing of propellants 151
Small engines 149
Small launchers 123
Small satellites 706
small thrusters 146
smart structures 228
social competence 742
soft landing 579
software 371
software development environment 464
software maintenance 464
software module 357
software tools 480, 760
Software Validation Facility (SVF) 733
soil condition 488
solar 271
solar array 239, 467
solar array panels 215
solar array power and battery energy 244
solar cell 237, 249
solar constant 36, 269
solar dynamic energy conversion process 239
Solar dynamic EPS systems 240
solar dynamic system 237
solar illumination 705
solar illumination and eclipse conditions 737
solar intensities 273
solar panels 541
Solar Physics 35

Solar Radiation 34, 36
solar radiation pressure 238, 338
solar sail 217
Solar Simulation 687
solar simulator 686, 687
solar spectrum 687
Solar System 575
solar thermal or electric propulsion 136
Solar Wind 38
solar zenith angle 270
solenoid valve 314
solid 154
solid elements 217
solid fuel boosters 136
solid fuel oxygen generator 424
Solid fuels 136
solid-state mass memories 532
solid-state power amplifiers 490, 558
Soret effect 601
sound pressure levels (SPL) 675
South Atlantic anomaly 42, 48
Soyuz 125
Soyuz rocket 5
Soyuz/Fregat 721
space 809
space age 1
Space Agency 824
Space astronomy 520
space conditions 682
Space Council 833
space craft control center 651
Space Debris 33, 51, 470, 832
space debris environment 101
space debris particles 100
space element 649
Space Environment 33, 43
space flight 17
space industry 3
Space Medicine 606
space probes 17, 29, 649
space segment 16, 652
Space Shuttle 12
Space Shuttle Orbiter 87
space simulation test facilities 686
Space Simulation Tests 682
space stations 9, 522
space system 657
Space Technology 554
space toilet 428
space transportation system 658
space vehicle 82
space-specific processes needed for product generation 770
space-wire data network 724
spacecraft 17, 464
spacecraft, control center, mission control software and spacecraft controller 465
Spacecraft Constellation Planning Facility (SCPF) 564
spacecraft operations manager 463
Spacecraft operations (platform and payload) 459
spacecraft status 471
spacecraft transport 115
spacecraft/payload functional tests 465
Spacelab 13, 522
spallation 108
SPAS 13
spatial densities of particles 105
spatial resolution 529
special checkout equipment 732
Special Environmental Tests 692
Special Optical Measurements of the Linear Expansion of Camera Structures 693
specialists 403
specific impulse 119, 155, 156, 304, 308, 316
specific impulse amounts 321
specific mission profiles 198
specifications 192, 785
specifications and work statements 769
specified operating temperature ranges 268
spectral power density 502
spectral resolution 525
spin stabilization 688
spin-stabilized Intelsat III satellite 319
spin-stabilized satellites 309
sporadic flux 100
Sports equipment 433
SPOT program 524
spread spectrum modulation 563
spreading codes 726
spring mechanisms 225
spring-driven hinges 216
Sputnik 1 519
sputtering 50
SSPC 267
Stability 333
Stability requirements 333
Stabilization 302

stable combustion process 179
stage equipment 153
stage hot firing test 198
stage separation 140, 196
stagnation pressure 151
standard deviation 72
standard formatted data units 472
standard Si cells 251
standardized telemetry and command packet formats 478
Stanton number 91
star patterns 342
state of the weather 523
State Responsibility 809
state space form 448
state transition matrix 71, 74
state vector 64
statement of work 772
states can agree among themselves 810
static and dynamic loads 195
static friction 233
static magnetic field configuration 691
station commander 404
statistical errors 74
steam pressure 49
Stefan–Boltzmann law 89
Step motors 228
step size 64
step track method 489
step track or sequential lobing 499
stereo method 530
sterilization process 440
stiffness 203, 207
Stirling process 240
storage properties 305
strategic cost calculation 793
stratosphere 39
streams 100
strength 203, 207
strength proof 220
Stribeck curve 233
strict (absolute) liability 831
String switching 246
structural components 195
structural mass 120
Structural Model 196
structural tests 220
structural–thermal model 203, 222
structure 25, 203
Structure and Thermal Model (STM) 715, 731
structure model (SM) 703
STS 131
Sublimation 49
suborbital rockets 17
subsystem specifications 277
success factors 741, 743
sudden overstraining 179
suitability for space applications 668
suitcase model 716
sum carrier-to-interference ratio 503
sum of the squared differences 71
Sun exclusion angle 343
Sun pointing 725
Sun-synchronous orbit 26, 76, 529
Supernova explosions 577
superorbital reentry 85
supervision 819
supplier audit 773
suppliers approved for space products 774
Support 776
support functions 798
surface charging 49, 212
surface coatings 271
surface heat flux 88
surface temperatures 432
surface tension 602
surface tension tanks 311
surfaces 283
SURV table 298
swath width 530
Swiveling mechanisms 149
synchronization and updating intervals 761
synchronized 726
synergies 670
synthetic aperture radar 248, 528
system 445, 658
system acceptance review 659
system analysis 658
system concepts 133
system data 726
System engineering 190
system engineering management 743
System engineers 20, 743
system experts 20
System integration 662, 714
system model 692
System Operations Validation 465
system requirements 277

System requirements review (SRR) 659, 749
system simulator 465
system validation test 465, 736
system verification 663
systematic data-driven processing sequences 535
Systematic errors 342
systems and operations of the ISS 408

T

TAI (Temps Atomique International) 568
tangential acceleration 449
tank fill rates 196
tanks 454
Target orbit 649
tasks 365
tasks of the project partners 742
tasks reacting in real time 374
TC uplink 563
teardrop tank 309
technical and methodological competence 742
technical competence 742
Technical Mitigation 759
technical qualification 818
technical specification 190
technical structure 532
technical water 429
technology demonstration 697
technology readiness level 121, 800
Teflon SSMs 290
telecommand 473, 485
telecommand system 491
telecommunication satellites 117
telecommunication services 820
telecommunications 485
telecommunications and meteorological 462
telemetry 21, 473, 485
telemetry and command system 478
telemetry, tracking and control 828
telemetry data 470, 471
telemetry data processing 465
telemetry packet types 461
telemetry system 147, 298, 492
telephone calls 405
telepresence 521
temperature 686
temperature control system 50
temperature control valve 423
temperature cycles 683
temperature cycling stress 684
temperature gradient 175
temperature predictions 282
Temperature regulation 295
temperature sensor 313
temperature variations 298
temporal variation of the orbital elements 61
temporarily stored 533
termination shock 584
terrestrial networks 552
terrestrial services 561
tertiary structures 214, 218
test adapter 673
Test Adapter Run 673
test campaign 670
test facilities and test centers 695
test matrix 669
test plan 669
test procedure 669
test readiness review 659
test sequence 670
Test Specification, Excitation Types, Theory 672
tethered satellite missions 45
Tethered satellite systems 238
*n*th-degree method 65
The antenna system 558
The European ground communication network 516
The laser retroreflector 559
the main constituents of a launcher 196
Thekaekara spectrum 687
thematic mapping 526
thermal analysis 294
thermal balance 46, 682, 733
thermal balance (TB) tests 278, 282, 683
thermal conditions 333
thermal conductivity 721
thermal contact conductance 274
thermal contact resistance 269
Thermal control 269
thermal control subsystem (TCS) 46, 275, 682
thermal cycling (TC) test 282, 684
thermal design 585, 588
thermal distortion (TD) tests 684
thermal efficiency 690
thermal emissivity 271
thermal energy 46, 238, 276
Thermal engineering 197
thermal environment of the launch site 150
Thermal insulation 294, 588
thermal mathematical model 152

Thermal microvibration tests 693
thermal noise 574
thermal operation limits 294
thermal protection systems 87, 97
thermal range 527
thermal resistance 274
thermal stress 692
Thermal Subsystem 18
Thermal System 686
thermal tests 285
thermal vacuum 737
thermal vacuum cycling (TV or TVC) tests 683
thermal vacuum test facilities 686
thermal vacuum testing 282
thermal/vacuum tests 682
Thermal–functional verification 197
thermionic energy converters 241
thermo-optical surfaces 269
thermodynamic conditions 151
thermodynamic cycle 238
thermoelastic distortion 686
thermoelastic distortion test 685
thermoelectric converters 237
thermoelectric energy converters 241
thermoelectric generator 588, 589
thermophysical properties 601
thermosolutal (buoyancy) 599
thermosphere 39
thin route traffic 552
third party is damaged 831
third-party liability insurance 818
three-axes-stabilized 346
three-axis actively stabilized satellite 302
three-axis attitude 339
three-axis stabilization 721
three-channel downconverter 488
three-dimensional coordinates and velocities 570
three-dimensional determination of local position 718
three-dimensional topographic maps 526
through-the-thickness inserts 212
Thrusters 317
thrust and braking maneuvers 75
thrust coefficient 155
thrust generation 119, 142
thrust levels 321
thrust vector control system 149
Thrusters for electric propulsion 306
tilt 120
Time and Cost Planning 753
time and frequency standard 492
time dependence of orbital motion 54
time displays 482
time line 405
Time Measurement 554
time signal 493
time synchronous signal 555
time update 73
time-delayed 484
timer code card 491
tip clearance 177
titanium-based alloys 205
TM downlink 563
TMR 379
TMTC 292
to position the propellants 141
tolerable magnetic moment 691
tools of a mission 506, 507
top-down approach 745, 750
top-down design 23
topocentric coordinate system 69
torques 348
torsion pendulum machines 681
torsion-spring-driven hinge 225
torsional brake 226
total impulse 313
total mass loss 211
total quality management 190
total velocity gain 119
trace gas sample line 421
trace gases 523
tracked 505
tracking 66, 486
tracking stations 188
training 508
Training and Simulations 465
training of special skills 408
training readiness review 410
training the staff 467
trajectory program 154
transfer mode 333
transfer of orbit and frequency usage rights 814
transfer of work package descriptions into processes 759
transfer orbit 587, 648
transfer segment 16
Transfer stages 136
Transient and shock loads 671

transients 179
TRANSIT 554
transition from a laminar to a turbulent boundary layer 86
Translation of the organization's quality policy 769
transmission 532
transmission of pictures and images 461
Transmit Power 542
transmitter position 559
transmitter unit 650
Transparency 744
transponder 548
transport 331
Transport Container 664
transport costs 115
transport systems 29, 437
transported horizontally 312
triangulation 554
tribological stresses 230
tribological systems 230
Tribosystems 230
trilateration 555
trilateration procedure 554
triol 432
troposphere 574
true anomaly 54
Tsiolkovsky equation 138
TT&C remote sites 562
turbosphere 40
turnkey contract 826
Twaron 206
two-axis measurements 345
two-failure 453
two-failure tolerance 450
two-line elements (TLE) 705
two-phase cooling loops 283
two-way propagation time measurements 563
TWT amplifier 490

U

UDMH 306
Uhuru 576
ultra rapid observed data 561
ultra rapid predicted data 561
ultrahigh-precision clock generator 728
Ultrastable oscillators 726
ultraviolet 36
ultraviolet X-rays and gamma radiation 34
umbilical 732
undamped oscillation 446
undamped system 179
underpressure valves 420
UNIDROIT 830
unified propulsion system 323
unintended emissions 688
uninterruptible power supply 477
unit dBi 497
United Nations 808
universal time 56
unlocking mechanisms 224
Unregulated Bus 247
unstable combustion process 179
unwanted emissions 689
upconverted 489
upconverter 491
Uplink 488
uplink and downlink 491
uplink modulator 491
uplink stations (ULS) 562
upper stages 136
Uranus 584
US export control 817
user centers 513
user ground centers 21
user interface 478
User segment 652
User Support and Operations Centers 506
user-driven processing sequences 535
utilization areas 2

V

V-bar 444
V-bar hopping 448
vacuum 47, 522, 705
Vacuum System 686
vacuum valves 421
validating operational procedures 467
value-added information products 535
valve 289, 322
Van Allen belt 211, 723
vapor flow 289
Variation of the Elements 61
variational equations 71
variations in signal runtime 484
Vega 124
vegetation 524
vehicle assembly building 185
vehicle equipment bay 145, 146

velocity 55, 155
velocity change 587
velocity gain 118
velocity requirement 120
velocity-compounded 177
vent line dump assembly 420
ventilation concept 421
venting 207
Venus 581, 585
Verification 193, 278, 357, 664, 769
verification and testing 669
Verification by analogy 192
verification control document (VCD) 755, 767, 792
verification matrix 195, 667
verification measures 192
verification of product requirements 791
Verification Planning 669
verification program 733
verified in orbit 697
verifies all interfaces 731
Very distant and old galaxies 577
very long baseline interferometry 523
vibration and thermal tests 699
vibration systems 673
vibration test 670
vibration test specifications 698
vibrations 260
vicinity of the target 443
video mode 491
Video System 476
virtual competition 797
viscosity 602
voice conferencing system 476
voltage operating point 255
Voltage variation at load changes 247
VOSDUKH 424
voter 377

W

Walker constellation 561
walking ellipse 447
walking machines 636
warm-up steps 298
warming 298
water as a propellant 307
water–copper heat pipes 288
water management system 426
water quality 430
water separator 431
water vapor desorption 424
water vapor regained 211
wavelength ranges 575
wax motor 225
Weapons of war 816
wear 231
Weather forecasts 520
Weather Observation 523
weather station 188
Web interface 481
weekly look-ahead plan 510
weight matrix 72
weighted average 64
Wernher von Braun 5
wheel unloading 356
Whipple shield 107
white paint 272
White Paper 833
Wideband Delphi Method 801
Wien's displacement law 271
winches 227
wind tunnel investigations 96
W. M. Kaula 61
work breakdown structure 752, 783, 798
work methodologies 406
work package description (WPD) 753, 783, 798
World Geodetic System 1984 (WGS84) 570

X

X-band 590
xenon 306
xenon peaks 687
XMM-Newton 576
X-ray range 520
X-rays 330

Y

yaw angle 336
Yo-Yo systems 229

Z

Zenit 721
Zenit-SL 126
zeoliths 424

Figures_main_chapters

Frontispiece: Computer generated image of the International Space Station.

Chapter 1: The first landing on the Moon during the Apollo-11 mission marks the most prominent event in human spaceflight. The image shows the lunar landing module "Eagle" returning to the Moon orbiter "Columbia". The photograph was taken by Michael Collins before docking on July 21st 1969, at 21:34:00 UT (Source: NASA).

Chapter 2: Artist view of the satellite TerraSAR-X in its space environment above Europe (Source: DLR).

Chapter 3: Ariane 5 ECA, flight V175, launches on March 11th 2007 from the European space port in Kourou (French Guiana) with the satellites Skynet 5A and Insat 4B (Source: ESA).

Chapter 4: Space shuttle Discovery (STS-121) on July 6th 2006 prior to docking at the International Space Station. The cargo bay holds the Leonardo Multipurpose Logistics Module (Source: NASA).

Chapter 5: The astronauts Robert L. Curbeam (left) and Christer Fuglesang (right) perform an Extra Vehicular Activity (EVA) at the International Space Station above New Zealand (Source: NASA).

Chapter 6: S-Band antenna with 30 m disc diameter at the DLR ground station complex in Weilheim, Bavaria (Source: DLR).

Chapter 7: Southern Bavaria, seen from the satellite Landsat. In the upper part the city of Munich is visible. The lakes "Ammersee" and "Starnberger See" are to be seen in the middle and the "Inn"-valley and the "Zillertal" in Austria in the lower part. The image was processed by the German Remote Sensing Data Center (DFD) at DLR (Source: DLR/University of Maryland, Global Land Cover Faicility – GLCF–Earth Sat).

Chapter 8: Students of the "Fachhochschule Aachen" with the pico satellite COMPASS ONE (Source: Fachhochschule Aachen).

Chapter 9: Artists' view of the Automated Transfer Vehicle (ATV) "Jules Verne" docked to the International Space Station (Source: NASA/ESA).

Last figure: Galaxy NGC 1672, 60 million light years away from Earth. The image has been taken by the Hubble space telescope (Source: NASA/ESA).

Handbook of Space Technology

Five decades after the launch of the first satellite – Sputnik 1 – in October 1957, this *Handbook of Space Technology* is intended to provide a comprehensive overview of this exciting field.

The handbook, with four-color printing throughout, provides not only students, engineers and scientists but also nonprofessionals who have a serious interest in space activities with detailed insights into the fascinating world of space technology.

The chapters

- Introduction (historical overview, space missions)
- Fundamentals (orbit mechanics, aerothermodynamics/reentry, space debris, etc.)
- Launch Vehicles (stage technology, propulsion systems, launch infrastructure, etc.)
- Spacecraft Subsystems (structure, energy supply, thermal regulation, attitude control, communication, etc.)
- Aspects of Human Space Flight (humans in space, life support systems, rendezvous and docking, etc.)
- Mission Operations (satellite operations, control center, ground station networks)
- Utilization of Space (earth observation, communication, navigation, astronomy, material sciences, space medicine, robotics, etc.)
- Configuration and Design of a Space System (mission concept, system concept, environmental simulation, system design, Galileo satellites, etc.)
- Management of Space Missions (project management, quality management, cost management, space law, etc.)

together with their 42 "subchapters" describe the processes and methodologies behind the development, construction, operation and utilization of space systems.

The individual chapters and subchapters of this handbook, written by leading experts from universities, research institutions and the space industry, enable the reader who wants information about a selected field to quickly gain a substantial overview. Those who wish to deepen their understanding of individual topics can refer to selected bibliographies as well as an extensive keyword index. The editors have taken care to structure the book in such a way that readers may also find it convenient to read it from cover to cover, thereby accessing a thorough presentation of the current status of space activities.

Editors:

Wilfried Ley

Professor of Space Technology, Fachhochschule Aachen

Klaus Wittmann

Director of Space Operations and Astronaut Training, German Aerospace Center (DLR), Oberpfaffenhofen

Willi Hallmann

Professor, formerly of Fachhochschule Aachen

Handbook of Space Technology Edited by Wilfried Ley, Klaus Wittmann and Willi Hallmann
© 2009 John Wiley & Sons, Ltd. ISBN: 978-0-470-69739-9