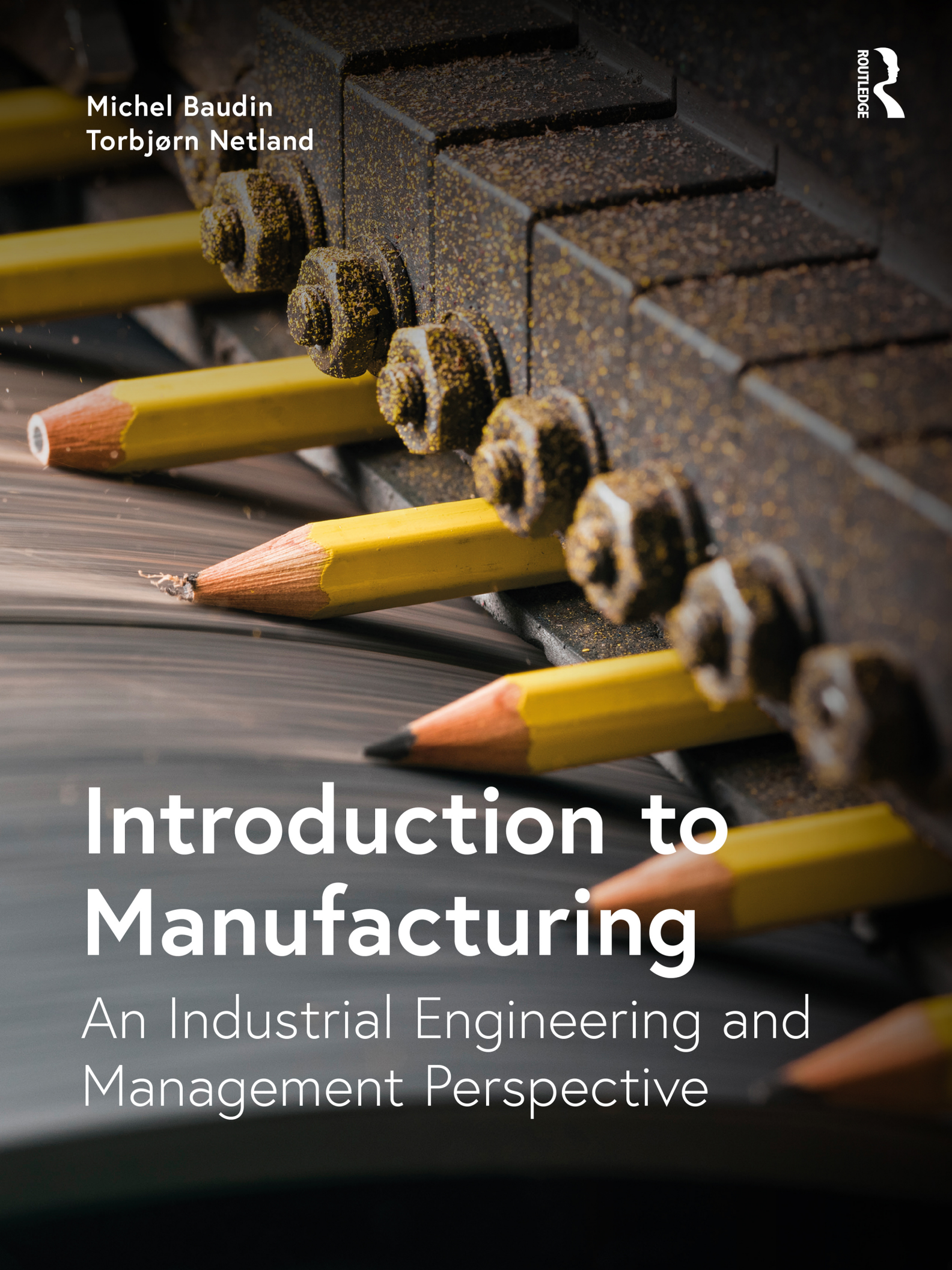


Michel Baudin
Torbjørn Netland

ROUTLEDGE

Introduction to Manufacturing

An Industrial Engineering and Management Perspective

Introduction to Manufacturing

This *Introduction to Manufacturing* focuses students on the issues that matter to practicing industrial engineers and managers. It offers a systems perspective on designing, managing, and improving manufacturing operations. On each topic, it covers the key issues, with pointers on where to dig deeper.

Unlike the many textbooks on operations management, supply chain management, and process technology, this book weaves together these threads as they interact in manufacturing. It has five parts:

1. *Getting to Know Manufacturing*: Fundamental concepts of manufacturing as an economic activity, from manufacturing strategy to forecasting market demand.
2. *Engineering the Factory*: Physical design of factories and processes, the necessary infrastructure and technology for manufacturing.
3. *Making Information Flow*: The "central nervous system" that triggers and responds to events occurring in production.
4. *Making Materials Flow*: The logistics of manufacturing, from materials handling inside the factory via warehousing to supply chain management.
5. *Enhancing Performance*: Managing manufacturing performance and methods to maintain and improve it, both in times of normal operations and emergencies.

Supported with rich illustrations and teaching aids, *Introduction to Manufacturing* is essential reading for industrial engineering and management students – of all ages and backgrounds – engaged in the vital task of making the things we all use.

Michel Baudin is an engineer, author, and consultant who graduated from Mines-ParisTech in 1977. His career has taken him from France to Japan, Germany, and finally to the US, where he lives in Palo Alto, California. Michel has taught courses in-house for clients, UC Berkeley, the University of Dayton, the Hong Kong Productivity Council, the University of Buckingham, and training companies in multiple countries. He has authored four books – *Manufacturing Systems Analysis* (1990), *Lean Assembly* (2002), *Lean Logistics* (2005), and *Working with Machines* (2007) – and written in refereed journals, professional magazines, and conference proceedings. Michel is a LinkedIn influencer and blogs at michelbaudin.com.

Torbjørn Netland is a professor and Head of the Chair of Production and Operations Management at ETH Zurich, Switzerland (pom.ethz.ch). He graduated from the Norwegian University of Science and Technology (NTNU) in 2006, and holds a doctorate from the same institution. Torbjørn has been a visiting scholar at Georgetown University and University of Cambridge. He is widely published in leading field journals, including *Management Science*, *Journal of Operations Management*, *Production and Operations Management*, *International Journal of Operations & Production Management*, and *MIT Sloan Management Review*, among others. He is the recipient of two Shingo Research Awards and numerous teaching awards. He blogs at better-operations.com.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Introduction to Manufacturing
An Industrial Engineering and Management Perspective

by Michel Baudin and Torbjørn Netland

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

Designed cover image: © Christopher Payne / Esto

First published 2023
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2023 Michel Baudin & Torbjørn H. Netland

The right of Michel Baudin & Torbjørn H. Netland to be identified as authors of this work has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data
Names: Baudin, Michel, author. | Netland, Torbjørn H., author.
Title: Introduction to manufacturing : an industrial engineering and management perspective / Michel Baudin and Torbjørn H. Netland.
Description: 1 Edition. | New York, NY : Routledge, 2023. | Includes bibliographical references and index. |
Identifiers: LCCN 2022034044 (print) | LCCN 2022034045 (ebook)
Subjects: LCSH: Manufacturing industries. | Human engineering. | Production management. | Manufacturing processes--Automation.
Classification: LCC HD9720.5 .B38 2023 (print) | LCC HD9720.5 (ebook) |
DDC 338.09--dc23/eng/20220729
LC record available at https://lccn.loc.gov/2022034044
LC ebook record available at https://lccn.loc.gov/2022034045

ISBN: 9780815361428 (hbk)
ISBN: 9780815363194 (pbk)
ISBN: 9781351110310 (ebk)

DOI: 10.4324/9781351110310

Typeset in ET Bembo and Europa
by the authors.

Publisher's Note
This book has been prepared from camera-ready copy provided by the authors.

Contents

Endorsements ix
Background xi
About the authors xii
Acknowledgments xiii
Guided Tour xv

Getting to Know Manufacturing 1

1 About Manufacturing 3
1.1. What is manufacturing? 4
1.2. Factory organization 13
1.3. The history of manufacturing 21
1.4. Role of manufacturing today 36

2 Manufacturing Strategy 43
2.1. Vision, mission, and goals 44
2.2. Strategy and tactics 46
2.3. Competitive priorities and capabilities 47
2.4. Strategy planning frameworks 50
2.5. Make or buy 58
2.6. Strategies to serve market demand 65
2.7. Focused factories versus capacity factories 67

3 Serving Market Demand 71
3.1. The art of forecasting 72
3.2. Demand volume, variety, and variability 76
3.3. Quantitative forecasting methods 82
3.4. Qualitative forecasting methods 88
3.5. Capacity planning 90
3.6. Takt time 104

Engineering the Factory 111

4 Process Design 113
4.1. General approach to process design 114
4.2. Drawing process maps 117
4.3. Classifying manufacturing processes 121
4.4. The Hayes-Wheelwright product-process matrix 125
4.5. Process capability 134
4.6. Machining cells 141
4.7. Operator Job Design and work combos 150

5 Assembly 155
5.1. Introduction to assembly processes 156
5.2. Assembly lines 159
5.3. Assembly line designs 165
5.4. Assembly cells 170
5.5. Assembly station design 176
5.6. Disassembly 187
5.7. Remanufacturing 189
5.8. Line balancing with Yamazumi charts 192

6 Automation 195
6.1. Background 196
6.2. Common automation methods 206
6.3. Principles of automation 213
6.4. Working with machines 216
6.5. The hidden factory 227

7 Layout 235
7.1. Macro and micro layout 236
7.2. Layout types 239
7.3. Role of flow in layout 243
7.4. Layout planning and design 250
7.5. Monuments 267
7.6. Layout decision-support methods 274

Making Information Flow 287

8 IT and OT in Manufacturing 289
8.1. The manufacturing IT/OT stack 290
8.2. Infrastructure 293
8.3. Data acquisition 307
8.4. Machine control 309
8.5. Supervisory process control 312
8.6. Operation support 314
8.7. Engineering and management support 319
8.8. Management of manufacturing IT/OT 323

9 Planning and Control 329
9.1. The planning challenge 330
9.2. Production planning overview 333
9.3. Leveled sequencing 336
9.4. Autonomous shop floor pull systems 350
9.5. Enterprise Resource Planning 368

10 Master Data Management 377
10.1. Introduction to master data 378
10.2. Nomenclature 381

10.3. Bills of materials 390
10.4. Equipment master data 410
10.5. Routings master data 412
10.6. Locations master data 414
10.7. External relations master data 415
10.8. Relationships between master data 415
10.9. Master data revision management 416

Making Materials Flow 421

11 Internal Logistics 423
11.1. What is the scope of internal logistics? 424
11.2. Water spiders and supermarkets 433
11.3. Pull systems 436
11.4. Part presentation 446
11.5. Logistics technology 454

12 External Logistics 469
12.1. Inbound and outbound logistics 470
12.2. Multimodal and intermodal transportation 471
12.3. Containerization 472
12.4. Incoterms 475
12.5. Packaging for external logistics 476
12.6. Container tracking 487
12.7. Third-party logistics 494
12.8. Supplier milk runs 499
12.9. Consolidation centers 504

13 Warehouse Management 513
13.1. The need for inventory 514
13.2. Warehouse planning and design 517
13.3. Manual versus automated storage and retrieval 535
13.4. Organizing the warehouse 543
13.5. Material stores inside the factory 549

14 Supply Chain Management 557
14.1. Supply chains and supply chain networks 558
14.2. Supply chain synchronization 564
14.3. Collaborative customer–supplier relationship 572
14.4. The outbound supply chain 580

Enhancing Performance 589

15 Managing Performance 591
15.1. Introduction to measurement 592
15.2. Metrics in manufacturing 600

15.3. Performance boards and meetings 615
15.4. Responding to emergencies 624

16 Improving Performance 639
16.1. The case for improvement 640
16.2. Kaizen 641
16.3. Improvement projects 652
16.4. Project management 658
16.5. Company-wide improvement models 665
16.6. Leading a transformation program 678

17 Improving Quality 685
17.1. Introduction to quality management 686
17.2. Inspection and testing 698
17.3. Data-driven quality improvement 705
17.4. Built-in quality control 715
17.5. Problem-solving frameworks and methods 722
17.6. When defects slip through 730

18 Maintenance 737
18.1. Introduction to maintenance 738
18.2. Organizing for maintenance 740
18.3. Approaches to maintenance improvement 749

Bibliography 765

Index 773

Endorsements

"What an incredibly timely book. Global events continue to starkly illustrate the critical role that manufacturing plays in all our lives – and its impact when things go wrong. Baudin and Netland's beautifully structured and accessible book gives an excellent introduction to this critical topic, helpfully framed by both industrial engineering and management perspectives."

— Tim Minshall, Dr John C Taylor Professor of Innovation
and Head of the Institute for Manufacturing, University of Cambridge

"Easy to read, yet rich in knowledge. This textbook is the ultimate introduction to anyone interested in a career in manufacturing."

— Gisela Lanza, Professor and Head of Production Systems
wbk Institute of Production Science, Karlsruhe Institute of Technology

"Given the every changing nature of manufacturing technologies, I love how the authors emphasize the flow of information and materials within and beyond the factory to provide an integrative approach to introduce manufacturing and management perspectives to industrial engineers."

— Timothy W. Simson, Paul Morrow Professor of Engineering Design & Manufacturing
College of Engineering, Penn State University

"Finally a textbook that describes manufacturing as what it is: an ensemble of processes, technology, and people. It is a terrific primer to manufacturing students and professionals."

— Li Zheng, Professor of Industrial Engineering
Department of Industrial Engineering, Tsinghua University

"What is really remarkable about this book is its sharp focus combined with coverage of a rich range of topics. The reader is captivated from the very first page, and with delightfully clear examples and crisp writing, the book takes the reader from a basic understanding of the fundamental concepts in manufacturing to highly sophisticated notions of the impact of manufacturing at micro and macro levels. This book is useful not just for students, but also for scholars."

— Kasra Ferdows, Professor of Global Manufacturing
McDonough School of Business, Georgetown University

"This state-of-the-art textbook on manufacturing addresses the challenges of today's factories. It is an essential manual for students and professionals."

— Jochen Deuse, Professor and Head of Institute of Production Systems, TU Dortmund
and Head of Centre for Advanced Manufacturing, University of Technology Sydney

"In this amazing book, Baudin and Netland present challenging technical concepts in an easy, accessible form that anyone can follow. Providing historical context in which industrial engineering concepts and operations management theories evolved, will be a game-changer for other texts in future. A must-read for academics and managers alike!"

— Rachna Shah, Professor of Supply Chain and Operations
Carlson School of Management, University of Minnesota

"This book introduces modern manufacturing systems using unique pedagogical contents with interesting real-life examples. Readers can easily grasp the essence of manufacturing and understand the core knowledge rapidly to be an effective learner."

— Jay Lee, L.W. Scott Alter Chair Professor in Advanced Manufacturing
and Founding Director of Industrial AI Center, University of Cincinnati

"For those embarking on a career in manufacturing this is a great place to start - learn from Toyota and others to harness the potential of the digital age."

— Daniel Jones, co-author of *The Machine that Changed the World* and other bestsellers
Founder and Chairman of the Lean Enterprise Academy

"Through an extremely timely new book, Baudin and Netland takes a recommendable fresh look at manufacturing in the 21st century. In the present complex age of sustainability, digitalization, and new roles for humans, the authors provide both industrialists and academics with a solid and accessible knowledge base for developing the future of manufacturing industry."

— Johan Stahre, Professor of Production Systems
Department of Industrial and Materials Science, Chalmers University of Technology

"Finally, an operations management textbook capable of supporting the current renaissance in manufacturing. Baudin and Netland provide a much-needed integration of learning from our modern experience with lean manufacturing, systems thinking, and information technology. I recommend this book for professors and students in technology-oriented disciplines that want a practical and systematic approach to making manufacturing operations work better."

— Eric O. Olsen, Professor of Industrial Technology
Orfalea College of Business, Cal Poly State University

"This comprehensive textbook covers manufacturing topics that matter the most to managers and engineers. This fills a gap in the existing set of textbooks. In addition to the comprehensive explanations, some great features of this textbook are the stop and think exercises, the additional reading lists, and the focus on improving manufacturing performance."

— Ken Snyder, Executive Director
Shingo Institute, Utah State University

Background

This book is the result of the meeting of two minds. As an engineer, a manager, and a consultant in manufacturing for 40 years, Michel Baudin has previously shared his experience in four books and a blog targeted at professionals. Torbjørn Netland – also an engineer by training and heart – is a professor of Production and Operations Management at ETH Zurich, a leading technical university. A decade ago, the authors met through blogging, and kept in touch, bonded by their shared passion for manufacturing. Michel had the idea to coauthor a textbook, and years later we are proud to release the product.

Between volumes of textbooks covering general operations and supply chain management on the one hand and manufacturing processes and technology on the other, we could not find any high-quality textbook that covered manufacturing from the perspective that matters the most for factory managers and industrial engineers. This was a disappointing discovery because many of the students we teach in higher education take on jobs that require in-depth knowledge about how to design and manage manufacturing activities. This book fills this gap. We were excited when we received very encouraging peer reviews on our book proposal from Routledge.

In this book, we put the spotlight on how to design, manage, and improve *factories*, which produce almost all goods and infrastructure we consume and use in our daily lives. Considering its importance to human prosperity and well-being, the science of manufacturing deserves to be taught much more broadly and deeply than it is today. We hope teachers, students, and practitioners enjoy reading this textbook on manufacturing, and find its content useful when designing, managing, and improving manufacturing operations.

You are reading the first edition of this textbook. While we have tried to be as complete and accurate as possible, there is always room for improvement. We would be glad to receive any feedback and comments you may have to this version.

About the authors

Michel Baudin is an engineer, author, and consultant who graduated from Mines-ParisTech in 1977. Moving from research to manufacturing, he felt like Buzz Lightyear in Toy Story, bemoaning "years of academy training, wasted." Once over the culture shock, however, he began to appreciate the relentless pragmatism of manufacturing, the camaraderie of factory people, and their ingenuity.

His career has taken him to Japan, Germany, and finally to the US, where he lives in Palo Alto, California. He has consulted for leading companies worldwide supplementing direct observation on the shop floor and stakeholder interviews with data mining on clients' systems, often identifying unknown patterns in product demand or manufacturing operations, leading to specific advice on both management and technology.

Michel has taught courses in-house for clients, and for UC Berkeley, the University of Dayton, the Hong Kong Productivity Council, the University of Buckingham, and training companies in multiple countries. He has authored four books: *Manufacturing Systems Analysis* (1990), *Lean Assembly* (2002), *Lean Logistics* (2005), and *Working with Machines* (2007). He has contributed chapters to the *Handbook of Supply Chain Management* (2006), *RFID Applications and Cases* (2006), and the *Routledge Companion to Lean Management* (2016). He has written in refereed journals, professional magazines, and conference proceedings. He is a LinkedIn influencer and has published 950 blog posts since 2011 at michelbaudin.com.

Torbjørn Netland is a professor of Production and Operations Management at ETH Zurich in Switzerland (pom.ethz.ch) and a co-founder of EthonAI, a tech startup offering AI solutions for quality management in manufacturing. He graduated from the Norwegian University of Science and Technology (NTNU), Norway, in 2006 and holds a doctorate from the same institution (2013). He has been a Fulbright visiting research fellow at Georgetown University's McDonough School of Business, a visiting scholar at the Institute for Manufacturing (IfM), University of Cambridge, and a Research Manager in SINTEF.

Torbjørn is currently a Member of the World Economic Forum's Global Future Council on Advanced Manufacturing and Value Chains, where he also serves on the expert panel for the Global Lighthouse Network. He is a Fellow of the European Academy for Industrial Management (AIM), Full Member of IFIP WG 5.7 Advances in Production Management Systems (APMS), President of POMS College of Operational Excellence, and a member of IEOM, IEEE, AOM, and EurOMA.

He is widely published in leading scientific journals in the area of manufacturing management, including *Management Science, Journal of Operations Management, Production and Operations Management*, *International Journal of Operations & Production Management*, *International Journal of Production Research*, and *MIT Sloan Management Review*, among others. He is the recipient of two Shingo Research Awards and several teaching awards. He blogs at better-operations.com.

Acknowledgments

This book is a result of hundreds of encounters we have made in manufacturing, research, and our personal lives. Collectively we want to thank "the Nerds" — Christoph Roser, Mark Warren, Franck Vermet, Hironori Hibino, Kai Lorentzen, Ralph Richter, Cécile Roche, and Peter Hines — for providing a community of practice to test and develop ideas. We also thank all the students we have been privileged to meet over the years, from whom we learn more than we teach. We thank Katalin Tesch at ETH Zurich for her invaluable help in editing the last versions of the book. Thanks are due also to individuals acknowledged throughout the book for letting us use some of their ideas.

Acknowledgments from Michel Baudin

Thanks to my wife Catherine for her patience and to my grown sons Thomas and Pierre for technical support. Much of what I am sharing I developed while working with manufacturing clients. In particular, I would like to thank Gregory Dumias, Walid El-Sayed, and their teams at AGCO, Dale Bayley at Parker Hannifin, Kevin Hop, Terry Mumaw, and Rick Regula at Honda. Kevin has since become a consulting colleague. I would also like to acknowledge Luc Delamotte at Thales.

I also learned much from my mentor Kei Abe in Japan, and many colleagues and associates. In the US, Crispin Vincenti-Brown and Hormoz Mogharei were valuable resources, and Katie Anderson, among other things, provided the daruma doll for the book. In France, I would like to acknowledge Christophe Caberlon, Gilbert Liégeois, and Radu Demetrescoux; in Russia, Marcel Zainashev, who is also a musician and a poet, Ernest Bashkardin, and Maria Samsonova, both of whom were much more than interpreters; in Spain, Jose Ignacio Erausquin, and the Asenta team.

From academia, I thank Carol Shaw from the University of Dayton, Charles Parks from Ohio University, Thomas Lee in his days at HKPC, and John Darlington at the University of Buckingham for engaging me in teaching. I also want to thank my coauthor, Torbjørn Netland, who is a brilliant academic. He is not only an original thinker and rigorous researcher but an outstanding teacher as well. I was thrilled when he agreed to partner with me on this project.

Acknowledgments from Torbjørn Netland

Thanks to Julia, Tobias, Nina, and Niklas for all your love and support. Professionally, I am indebted to the many manufacturers that I and my team at ETH Zurich have been fortunate to work extensively with over the years, including ABB, Aker, Dätwyler, Geberit, GKN, Hitachi, Hydro, IKEA, Jotun, Kongsberg, Planted, Siemens, Volvo Group, and many more. Thanks also to the over 100 graduate students who have written their Master's theses in various manufacturing companies under my supervision.

In academia, I have been fortunate to be taught, challenged, and mentored by outstanding individuals. I'd like to highlight Kasra Ferdows, Jag Srai, Jason Schloetzer, Stephan Wagner, Rachna Shah, Arild Aspelund, Daryl Powell, numerous ex-colleagues at SINTEF and NTNU, the many co-members of the WEF Global Future Council, and all employees and alumni of the Chair of Production and Operations Management, ETH Zurich. Last but not least, this book could not have been written without the passion and experience of my coauthor Michel Baudin, who is the most knowledgeable person I know of in the area of manufacturing.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Guided Tour

This *Introduction to Manufacturing* focuses on the strategic, tactical, and operational issues related to the engineering and management of manufacturing operations in factories, and the immediate links to suppliers and customers. It provides details on how operations can and should be designed and organized in a factory, and on the management of technology and people.

The parts

This book is divided into five parts:

1. **Getting to Know Manufacturing**: An introduction to the fundamental concepts of manufacturing as an economic activity, including serving market demand and manufacturing strategy.
2. **Engineering the Factory**: The physical design of factories and processes, explaining how to establish the necessary infrastructure and technology for manufacturing.
3. **Making Information Flow**: The information flows are the "central nervous system" that triggers and responds to events occurring in production.
4. **Making Materials Flow**: The logistics of manufacturing, ranging from materials handling inside the factory to supply chain management.
5. **Enhancing Performance**: The improvement of manufacturing activities, including the monitoring of performance and methods to improve it, both in times of normal operations and emergencies.

Part 1, *Getting to Know Manufacturing*, considers manufacturing from the outside and from the boardroom. The subject has fuzzy boundaries but, for the purpose of this book, we sharpen them to clarify what we are discussing in *About Manufacturing* (Chapter 1). We describe milestones in the development of manufacturing, including the role it plays in both emerging and advanced economies. Having set this context, we describe various methods for setting *Manufacturing Strategy* (Chapter 2), through brief overviews and examples of companies from different industries. The next step is to characterize the market demand for the products to be manufactured, the manufacturing capacity needed to meet it, and the requirements to suppliers. It is covered in *Serving Market Demand* (Chapter 3).

In Part 2, *Engineering the Factory*, we enter the factory and focus on the engineering challenges of making information and materials flow through all the operations performed in-house. First, we look at *Process Design* (Chapter 4). *Assembly* (Chapter 5) is commonly the final process in discrete manufacturing, and, since it tends to be labor intensive it deserves a separate chapter. Our perspective on *Automation* (Chapter 6) is rooted in Toyota's jidoka, which considers production as a human-machine system, as opposed to just a machine system. Then we look at *Layout* (Chapter 7), which is the physical organization of processes – networks of operations turning input materials into output products.

Part 3, *Making Information Flow,* is about information, which is and always has been as vital to manufacturing as materials. Without it, operators don't know what to work on, how to do and validate the work, or how to respond to anomalies. Programmability is the most sweeping change to have occurred in the way humans interact with machines, and it has transformed the work of managers, engineers, maintenance technicians, and production operators. We discuss the impact of *Operational Technology* (OT) and *Information Technology* (IT) (Chapter 8), from what happens inside production machines to high-level support, and then go deeper into the specific issues related to *Planning and Control* (Chapter 9) and *Master Data Management* (Chapter 10).

In Part 4, *Making Materials Flow*, we start with *Internal Logistics* (Chapter 11) – the work inside the factory walls – and move on to *External Logistics* (Chapter 12) – which happens outside and involves collaboration with suppliers and possibly third-party logistics services. Parts and products usually have to be stored, which requires skills in *Warehouse Management* (Chapter 13). As suppliers are independent business entities, *Supply Chain Management* (Chapter 14), seen from a manufacturer's perspective, encompasses the design and analysis of a supply chain, techniques to reduce or eliminate the wild swings known as the "bullwhip effect," and supplier development.

Part 5, *Enhancing Performance* is about what manufacturers should do once the concepts and methods from Parts 1 to 4 are roughly established. The flows of information and materials and the work of people should be continuously improved to enhance the sustainable competitiveness of a manufacturing company. This, in turn, begs the question of defining relevant, understandable, and actionable performance metrics and using them in a way that avoids counterproductive gaming. This is *Managing Performance* (Chapter 15), which must be done both during normal operations and emergencies. The experience of the past decades in Japan and elsewhere has yielded a number of approaches to the planning and management of improvement activities of varying scopes and durations, which constitute a body of knowledge on *Improving Performance* (Chapter 16). It suggests that *Improving Quality* (Chapter 17) should receive special attention, which it does in this book. Finally, besides programming machines, maintaining them is the other task that will ultimately remain for people. *Maintenance* (Chapter 18) is the last chapter of this book.

We have integrated a number of pedagogical elements that will support readers in their learning journey. For example, as outlined above, the book is given a clear structure that advances the learner's knowledge in a logical manner. Each chapter includes a list of *Learning Objectives*, *Stop-and-Think* questions, and rich *Box* examples. We have put emphasis on illustrations and intentionally included many photographs of real manufacturing activities to help readers get an understanding of the diversity and complexities involved. Theoretical ideas are presented throughout the book and explained with real-life examples. Key references to theoretical and foundational work is given in footnotes. Each chapter ends with a *Summary*, *Test yourself* questions, and a *Further Reading* list. These features are all intended to guide and help the reader in their learning journeys. We hope that this textbook will become an essential resource for students of industrial engineering and operations management – in all stages of their careers.

Editorial choices

Manufacturing is ripe with business jargon. For example, "six sigma," "Total Quality Management," "Total Productive Maintenance," "lean," "agile," and "Industry 4.0," among many others, have developed from specific ideas and concepts into brands. People working in or with factories come up with new ideas that enhance manufacturing performance. Consultants and academics then generalize these ideas and help diffuse them, branding them as needed to raise attention. Over time, the core ideas join the mainstream manufacturing body of knowledge and the brands fade away as they are replaced by new ones. In this textbook, we focus on the *body of knowledge*, not brands.

In manufacturing, Japanese firms – and particularly Toyota Motor Manufacturing – have contributed much to the body of knowledge. Many Japanese words have become part of everyday factory language all over the world. We include the original Japanese spelling of these words, as in "jidoka" (自働化), which is for the benefit of students in Japan and in the parts of Asia that use Chinese characters.

The manufacturing literature emphasizes standardization, yet, at the most basic level of weights and measures, several systems are still in use. The International System of Units (SI), the latest version of the metric system, is used in nearly every country in science, technology, industry, and commerce. The notable exception is the US, which uses its own system based on the old British system. In this book, we present case studies from multiple countries and have kept for each the system of units in which it was originally formulated. This helps students develop sensitivity to different units and the associated problems.

We have ensured the rights to use the photos and illustrations in this book. Most of them are our own. To improve readability, we have not included credits to our own work, nor to photographs bought from a stock photo service that doesn't require it (Envato Elements). We include credits to photos and illustrations from other sources.

We have capitalized the first letters of words or concepts that can otherwise be ambiguous, for example, "Quality Circle" or "Total Productive Maintenance." Types of IT systems have also been capitalized, as in "Manufacturing Execution System" or "Virtual Reality." When we refer to a specific department within a manufacturing organization, we use capitalized first letters as in "Assembly" or "Sales." But to avoid excessive capitalization, we write unambiguous concepts like "lean," "last-in-first-out," and "just-in-time" with lowercase letters. We also write all Japanese words with lowercase letters, as in "kanban" or "jidoka." We use *italic* to emphasize words. Although we introduce abbreviations, we often use or repeat the unabbreviated words to make the text easier to follow.

We name inventors and the problems they were trying to solve in the main chapter text. To increase readability, we use footnotes to acknowledge authors of books and articles in which original ideas and concepts are presented, and summarize them in a Bibliography list in the book's back matter. The references are in APA format. We encourage students to also check out the many great resources listed in the Further Reading section of each chapter. And don't forget; in manufacturing, there is no greater teacher than the gemba!



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Part 1

Getting to Know Manufacturing

DOI:10.4324/9781351110310-1



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

1 About Manufacturing

Manufacturing takes place in factories. A factory is the adventurous setting in the famous novel *Charlie and the Chocolate Factory*, where Roald Dahl (1964) wrote, "I, Willy Wonka, have decided to allow five children – just five, mind you, and no more – to visit my factory this year." But, "The Factory" is also the name of a location in the TV series *Colony* (2016). Little is known about "The Factory" in Colony, except that it is the last place you want to be, it is a slave labor facility located on the Moon, and nobody has ever returned from it… Outsiders depict factories as both the source of "good manufacturing jobs" and "the world's most hated workplace." We offer a more factual view of manufacturing in the 21st century than Willy Wonka's enchanting chocolate factory and the spine-chilling sweatshop in Colony.

For the hundreds of millions of people who work in factories all over the world, making every physical object we use and consume, it is often the best economic opportunity available. Production work can be tedious, but workers still experience pride in workmanship, particularly when the products have a reputation for high quality. In addition, the factory is a place of intense teamwork that fosters a camaraderie among coworkers that is uncommon in other workplaces. The work of designing and operating factories requires a combination of technical and managerial skills that is best grounded in a fascination with how things are made. It is often a calling, not a pursuit based on economic necessity or incentives.

Manufacturing contributes to society, not only through its output, but also through the impact of factories on their social and physical environment. Whether or not they measure up, it is the responsibility of owners and managers to make manufacturing jobs foster economic growth, provide career opportunities, and ensure that factories do not pollute or harm people. This chapter provides you the necessary background knowledge to understand what manufacturing is about.

Learning objectives

After working through this chapter, you will be able to do the following:

1. Tell manufacturing apart from other business activities.
2. Quantify the importance of manufacturing in the economy.
3. Map a manufacturing activity as a system of processes.
4. Explain the concepts of effectiveness and efficiency and their relation.
5. Use the MANUFACTURES mnemonic to describe components of manufacturing.
6. Design an organizational structure for a typical factory.
7. Explain the role of operators, support staff, and managers.
8. Explain how manufacturing developed through history.

DOI:10.4324/9781351110310-2

1.1. What is manufacturing?

Many operations can be described as turning some inputs into outputs, but not all do so by manufacturing. The physical nature of the goods is essential to manufacturing. It drives the technical, human and economic organization of the work. The emphasis on physical materials and goods does not imply that digital and analog information or service processes are of less relevance. It just means that, unless the goods are physical, it's not manufacturing.

Here are some examples and counterexamples of what manufacturing is and is not:

- Extracting metals or rare earths from ores is manufacturing, and so is building mining equipment. Yet mining itself, on the other hand, is not manufacturing. It's about getting materials out of the ground, wherever they happen to be, which is not the same as transforming them. Mining has been around longer than manufacturing and has contributed many ideas used in manufacturing. For example, the concept that a manager's job is to plan, organize, lead, and control was developed by mining executive Henri Fayol.
- Making 6 million memory chips in a factory per month is manufacturing, but making a prototype of a new chip in the factory lab is not.
- Cooking, packaging, and freezing meals for sale to supermarkets is manufacturing, but operating a restaurant is not. In restaurants, the business is food service, not production. Cooking a meal at home is not manufacturing because family members do not usually pay.
- Printing and binding textbooks for distribution through bookstores is manufacturing. Preparing and selling e-books is not, because it does not involve any materials. Producing the e-book readers is manufacturing.
- Replicating software on CDs, packaging and shrink-wrapping them is manufacturing, but selling software through downloads or as a service is not because no materials are transformed.
- Sculptors and painters produce physical goods, but their art is not manufacturing because each piece is unique and intended to be irreproducible. Manufacturing is usually repetitive, but it does not have to be. For example, shipbuilders or specialized machinery manufacturers may produce only one-of-a-kind products. In these examples, however, processes are repeatable. A manufacturer may be able to make a fully custom product, but can replicate it exactly if the need arises. On the other hand, copies or reproductions of works of art are never exact and usually do not have the same value as originals.

> **Definition**:
> Manufacturing is the transformation of physical materials into physical goods using repeatable processes.

Manufacturing takes place in factories or plants, which are buildings or complexes of buildings equipped with machines, tools, and information that people use to transform materials or assemble parts into products. Even the most advanced factories are not unattended automatic systems, but systems that function based on interactions between humans and machines. Factories can be found in many shapes and locations around the world. Figure 1.1 shows one in Brooklyn, New York.



Figure 1.1 Manufacturing takes place in factories, like this one in Brooklyn, New York

Manufacturing is most commonly done with the goal of directly generating profits from the sale of goods to customers. The relationship of manufacturing to profits is, however, often indirect. For example, firm-internal production does not directly result in sales to external customers. This includes making components used for assembly elsewhere in the company or making tools that the company uses in performing services. Furthermore, some defense goods are made in government armories with no profit motive. Such factories are "cost centers," to which the company allocates a budget to make the products and the factories are measured on their ability to meet it.

People are at the center of manufacturing. Over time, the number and the roles of people in a factory change, but they remain essential. For example, Nippon Steel's Yawata Steel Works, in Kitakyushu, Japan, produces as much steel today with less than 3,000 employees as it did in 1964 with more than 40,000. The work these 3,000 people do today is obviously different from the work the 40,000 did five decades before. Improvements in work methods, processes, control systems, and information technology have all contributed to this transition. In the end, however, the steel works remain a human–machine system. As we will see in later chapters, while technological advancement changes what people do, it does not eliminate the need for them.

1.1.1. Input-process-output model

On first encounter, the work of a factory may appear dauntingly complex. See, for example, the petrochemical industrial zone shown in Figure 1.2. What it boils down to, however, is that manufacturing is about processing input materials into outputs – effectively and efficiently. This insight is valid for each process in a factory. Hence, even the most complex manufacturing task can be broken down into processes and subprocesses.

Box 1.1 **Taxonomy of manufacturing by the United Nations**

Several governmental and non-governmental institutions offer standard definitions and taxonomies of what constitutes manufacturing and what not. All of them have their strengths and limitations. One of the most accepted classification schemes is the International Standard Industrial Classification (ISIC) of All Economic Activities by the United Nations.[1] In the version from 2008, ISIC defines manufacturing by the ISIC Section C divisions 10–33 listed here.

Section C Manufacturing

Division 10	Food products
Division 11	Beverages
Division 12	Tobacco products
Division 13	Textiles
Division 14	Wearing apparel
Division 15	Leather and related products
Division 16	Wood and products of wood and cork, except furniture; articles of straw and plaiting materials
Division 17	Paper and paper products
Division 18	Printing and reproduction of recorded media
Division 19	Coke and refined petroleum products
Division 20	Chemicals and chemical products
Division 21	Pharmaceuticals, medicinal chemical and botanical products
Division 22	Rubber and plastics products
Division 23	Other non-metallic mineral products
Division 24	Basic metals
Division 25	Fabricated metal products, except machinery and equipment
Division 26	Computers, electronic and optical products
Division 27	Electrical equipment
Division 28	Machinery and equipment not elsewhere classified
Division 29	Motor vehicles, trailers and semi-trailers
Division 30	Other transport equipment
Division 31	Furniture
Division 32	Other manufacturing (jewelry, musical instruments, sports goods, games and toys, medical and dental instruments, other manufacturing not elsewhere classified)
Division 33	Repair and installation of machinery and equipment

1 United Nations. (1990). Standard industrial classification of all economic activities, Revision 3, Series M, No. 4, Rev. 4. URL: https://unstats. un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf



Figure 1.2 Apparent complexity in a petrochemical industrial zone

There are many processes in a factory and many levels of processes. To understand how factories operate, it is useful to map processes at different levels using the input-process-output (IPO) model (Figure 1.3), where the output of a process is the input to one or more subsequent processes. The IPO model can be used to map and understand operations on all levels. On a factory level, we can use it to describe how a steel mill uses labor, ovens, machines, and energy to turn iron ore into steel products in a smelting and cooling process. The process can be broken down and detailed as we move from the factory to the shop floor level. For example, the overall factory process can be broken into several sub-processes: purchasing and logistics, iron making, steel making, casting, and rolling, dependent on the actual steel mill. Each of these processes have their own inputs and outputs, and each process can be further broken down into more sub-processes.

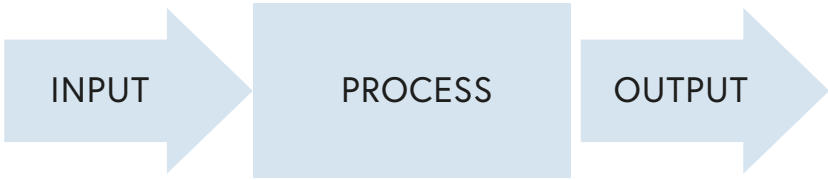


Figure 1.3 Input-process-output model

The IPO model implies a sequence, and the process breaks down into operations such that the output of each is the input to the next one. IPO models are used to map information flows as well as material flows, with one vital difference in the meaning of flow arrows. Information outputs are read by the next process. It is a replication operation, after which the information is still available for other uses in unlimited quantities. Materials, on the other hand, move to the next process and are no longer available, meaning that quantities must be accounted for.

Stop and Think 1.1
Can you use the IPO model to describe how you make tea or coffee? Is there any process you can think of where the IPO model does not apply?

Strictly speaking, not all processes in manufacturing take an input and process it into a physical output. Some processes affect a state change in a system rather than convert an input to an output. For example, repair operations in maintenance respond to an alert about a machine going into a "down" state, and the result of the repair is the return to "ready" status. Other examples are the processes of managing technical data or starting up a machine. The same distinction exists in control theory, classical control theory involves feedback where the process is a black box known only by the relationship of output to input. In modern control theory, the model is enriched with state variables.

Manufacturers often send old machines to overhaul or refurbishment factories that take them apart, replace worn or broken components, reassemble, and test them. The input-process-output model is a natural fit for the refurbishment factory because you actually have a flow of old machines coming in, being processed and being shipped out. The same model is a force-fit when you apply it to in-house maintenance, where machines don't move and teams of technicians are dispatched to them with all their equipment. You can apply the input-process-output model to work order information flows but not always to service work itself.

1.1.2. The two EFFs: Effectiveness and efficiency

A fundamental concept in manufacturing is the two "EFFs:" Effectiveness is about the "what" and efficiency about the "how." Being effective means getting the right things done. An effective process produces the intended outcome. Hence, it is measured during the output stage of the IPO model.

Regarding effectiveness, it doesn't really matter how the product was produced, as long as it fulfills the requirement of the customer or next process down the line. For example, if you order one standard pencil and a factory uses 1,000 trees to make just one pencil, it would be effective but not efficient. A perfectly efficient factory would use no more wood and carbon than actually goes into your pencil.

> **Definition**:
> Effectiveness means doing the right things, efficiency doing things right.

Efficiency is measured during the process stage of the IPO model. It is a ratio. For example, if all the heat provided to a steam engine were converted to mechanical energy, its efficiency would be 100%; if none of it were, it would be 0%. There is no way efficiency could be outside of this range. Likewise, if all the resources provided to a manufacturing process went directly into the sold products, its efficiency would be 100%. The manufacturing equivalent of friction and heat losses in the engine is overproduction, waiting, handling, moving, storing and retrieving, and making defective goods.

A perfectly effective and efficient factory is a vision. We can imagine it but real factories are always approximations. "Optimal" factories do not exist. The operations of real factories can always be improved, until disruption from new technology resets the improvement process at a higher level. Note that effectiveness should always take priority over efficiency. Being efficient at doing the wrong things is perhaps the biggest waste and not worth pursuing.

1.1.3. The four Ms

Another basic concept for process understanding is the four Ms. At a tactical level, problem-solving teams have been taught for decades to look for root causes under the four headings of "Man, Machine, Methods, and Materials." While it has been a useful guideline, we must adjust the wording today, because women should not be subsumed under the "Man" label. The four Ms now stand for "huMan, Machine, Methods, and Materials."

All the four Ms must be present at the same time for manufacturing to happen. Consider a welding process in a factory making exhaust pipes (Figure 1.4). With any of the following inputs missing, there will be no production:

- **Human**: Welder or human controlling or programming a welding robot
- **Machine**: Welding tool, protective equipment
- **Materials**: Steel pipes, welding electrodes, energy
- **Method**: Welding instructions or know-how, production plan.



Figure 1.4 The four Ms are needed for any process to run (Credit: kasto/123RF.com)

Even in fully automated processes, where workers apparently are not physically present, some humans will be indirectly involved, as programmers or troubleshooters. The method can be installed in machine programs or described in job instructions that are codified knowledge, or embedded in people's skills and know-how – that is, tacit knowledge.

While the four Ms concept may seem trivial, it is not. In a factory, every single process taking place at any time needs the four Ms *simultaneously* present to function. That quickly amounts to thousands of flows that must be coordinated.

Stop and Think 1.2
What are the 4Ms when you cook dinner and brush your teeth? Can you cook dinner or brush your teeth with one of the 4Ms missing?

1.1.4. The MANUFACTURES list

Lists are useful as memory joggers. However, short lists are often incomplete and long lists are difficult to remember. In practice, lists are a one-way filter. If you omit an item from the list, you know that your work is incomplete. On the other hand, checking off every item does not guarantee that it is complete. The following list uses MANUFACTURES as a mnemonic, where each letter stands for a topic that must be considered when designing or running manufacturing operations. The order of the topics is meant to fit the acronym and is not otherwise meaningful.

MANUFACTURES is short for Management, Agents, Nature, Users, Funds, Approach, Context, Technology, Utilities, Raw materials, Economics, and Suppliers. The following list explains each point in more detail.

- MANAGEMENT: The tasks of management in general are to plan, organize, lead, and control. This applies to manufacturing, in specific ways, not only to the physical flows of production but also to the required support activities like maintenance, purchasing, warehouse operations, quality assurance, human resources, accounting, etc.
- AGENTS: Agents are all individual stakeholders. A factory provides livelihoods to operators, technicians, engineers, and managers, and profits to its owners. It also affects the community of people who are dependent on the products or the employees as family members or service providers, or affected by its activities through noise, traffic, or other nuisances.
- NATURE: Regardless of management ambition, vision, or desire, the laws of physics, chemistry, or biology never cease to apply on the shop floor. Understanding these laws is vital to the production of a consistent product from materials with characteristics that vary, because they are extracted from the ground, are agricultural crops, are recycled, or simply come from multiple sources. Furthermore, manufacturing should not harm the environment.
- USERS: This includes the organizations or individuals who buy the products and the end-users. They are not always the same. Household appliances are usually bought by the people who use them, but from a retailer, not the manufacturer. Consequently, information about the products' popularity and quality is aggregated and delayed in the distribution network before reaching the manufacturer.
- FUNDS: This includes the investment in factories and production lines, operational expenses, and revenues from sales. The language of money is spoken at the boundary between the factory and its business environment. Activities inside the plant, on the other hand, are described in the language of things, like pallets, pieces, tons, or information, like plans, lead times, or yields. A challenge is to make local operational decisions in the language of things and information that translate into the language of money. This is the job of accounting and it is not easy.
- APPROACH: The approach describes the principles behind the design and methods used in operations. Mass production, as developed in the 1910s at Ford and refined at GM in the 1920s is one approach. The Toyota Production System (TPS), or lean manufacturing, is another. In broad strokes, they appear similar; in detail, they are not. It can be helpful to think of the approach as a recipe for preparing food or typical assembly instructions included in furniture or toys products. But note that written descriptions of the approach will never be complete descriptions of the process. For example, a typical recipe for making lasagna does not tell you where to cook, nor does it tell you where to find a spoon or how to move your arms when mixing ingredients for the sauce. The approach should be derived from the type of manufacturing processes.
- CONTEXT: The context describes the political, economic, and social environment in which manufacturing takes place. Local governments routinely offer tax exemptions, or free facilities, improve the road network, and provide other benefits to companies in exchange for locating a

factory in their community, creating jobs, and expanding the tax base. Conversely, the closure of a factory can be detrimental for a local economy.

- TECHNOLOGY: Technology covers the processes and equipment used in manufacturing. Any company with money can buy equipment, but it doesn't make it into a manufacturer any more than the possession of a musical instrument makes you a musician. In addition to the equipment, a manufacturer possesses the skills to use it to produce a specific result.
- UTILITIES: Utilities include availability and cost for the energy needed to run facilities and equipment, water that may be consumed in the processes, and waste disposal. Different manufacturing processes have different requirements. For example, access to affordable and reliable energy is a decision criterion for metal smelting plants.
- RAW MATERIALS: "Raw materials" here refer to all the materials that end up in the product, regardless of whether they are themselves the output of an upstream manufacturing process. The challenge is to ensure that the right materials are available at the right time in the right quantities. The materials must also be delivered at the right locations inside the factory and be presented for easy picking by operators or loading into machines.
- ECONOMICS: A factory must provide economic value to the company, but that can take many forms. The most straightforward motive is to profit from the sales of the goods but it is by no means the only one. Sometimes, the company prefers to make components in-house, even if it is more expensive than outsourcing, in order to have exclusive access to the plant's capacity and prevent know-how from leaking to competitors. Other companies set up a factory in the heart of an industry concentration, where they can grow their know-how by tapping into the local talent pool. Moreover, in a given factory, the objectives for generating economic returns change over time.
- SUPPLIERS: Suppliers are other companies that provide inputs to the manufacturing process. A manufacturer buys from them but does not have formal authority over them. As a customer, the manufacturer can influence its suppliers, but it is an order of magnitude more difficult than working with internal departments.

Please note that any attempt to summarize complex socio-technical concepts as simple rules to remember will have flaws. The idea with MANUFACTURES is not to present a complete theory of manufacturing, but to provide a rule of thumb that offers practical utility.

> **Stop and Think 1.3**
> Consider the example of a strawberry jam manufacturer. Can you give a few examples for each of the factors in MANUFACTURES?

1.1.5. Systems thinking

Although we try to simplify with IPO models and the MANUFACTURES mnemonic, to help us understand and manage manufacturing operations, a factory is in reality a complex system with interlinked social-, economic-, and technical components. It is the workplace of many people, it exists to generate a return to its investors, and it hosts a range of different technologies. Without an overall systems perspective, sub-optimization and misguided beliefs are sure to exist. The ancient parable from India about the elephant and six blind men provides a useful analogy (Figure 1.5).

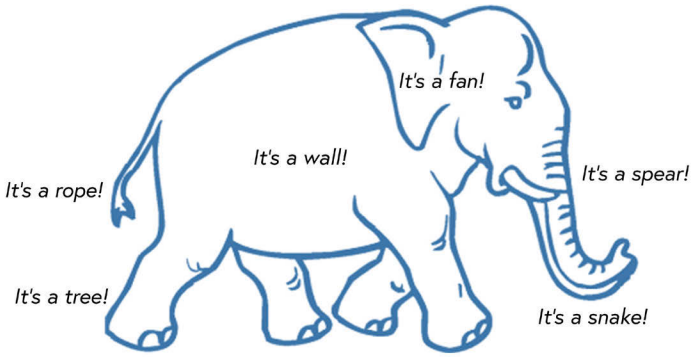


Figure 1.5 An elephant according to six blind men

Six blind men encounter an elephant for the first time and each one tries to describe the elephant based on a small piece of information from different perspectives. As shown in Figure 1.5, one touches the tail and thinks it feels like a rope. Another one touches a leg and suggests it is a tree stem. A third blind man hits the side of the elephant and experiences it as a wall. A fourth one touches the trunk and suggests it is a snake. One touches an ear and says it has to be a fan. The sixth blind man touches a tusk and describes it as a spear. Individually, all of them are far from a good description of an elephant. Together, they start to get close.

The lesson from this anecdote is that factories must be understood as systems, which successful manufacturers have known for a long time. At MIT in the 1950s and 1960s, Jay W. Forrester undertook to develop mathematical models of manufacturing activities and analyze them through computer simulations. He called it "Industrial Dynamics."[2] Today known as "systems thinking," it helps us explain how independent yet interlinked objects affect the processes and performance of a system. In manufacturing, W. Edwards Deming argued that the job of a manager is to work to improve the whole, not the parts.[3] Figure 1.6 shows how the Dragon capsule is lifted to be placed atop its cargo ring inside a processing hangar at Cape Canaveral Air Force Station in Florida, USA.



Figure 1.6 NASA uses systems thinking (Credit: NASA/Kim Shiflett, 2011)

2 Forrester, J. W. (1961). *Industrial dynamics*. M.I.T. Press.

3 Deming, W. E. (1986). *Out of the crisis: Quality, productivity and competitive position*. Cambridge University Press.

All elements of a space rocket are carefully planned with a view to the total system. A system is a set of elements that stand in relationship with each other and interact. Further, a system is defined by its boundaries that separate it from the environment.

A factory is a system that consists of humans, machines, materials, and so on, that all stand in various relationships to each other. These elements can be separated into soft — humans and information — and hard — machines and products. As we cannot fully predict the influences of one change in the element to all other elements, a factory system can be defined as being complex. With a complex system, the outcome of a change is not *a priori* predictable because of the dynamic relationships between the different elements that result in unexpected effects.

What does systems thinking mean in practice? Consider the example of increasing the output of a production line for computers. There are a few things you can do at the actual line; for example, installing new assembly robots, rebalancing the line, speeding up the line, etc. What if the shorter throughput time is lost in the packaging line? What if assembly workers feel more stressed and make mistakes? Are we sure that incoming logistics is able to match the new pace? When we design, manage, and improve manufacturing operations we must think of the factory and its supply chain as a system.

1.2. Factory organization

A factory is, among other things, a community of people who must be organized. People must take different roles and jobs to fulfill the objective of the factory, which is to make money by making products for a market. Different departments grow different cultures – even within the same factory. Communication barriers and silo-thinking curtail manufacturing performance. Hence, organizing the factory employees well can be a considerable advantage.

1.2.1. Organization charts

Organizations are often presented through an *organization chart*, also known as an organigram. An organization chart is a visual representation of the organizational structure. There is no standard on how to draw the organization chart other than placing departments in the order of their rank and show reporting relationships with lines, and indirect relationships with dotted lines. It is unusual to find the exact same organization chart in two different factories.

A limitation with an organization chart is that they put the focus exclusively on management reporting relationships. In particular, such charts fail to show the routine exchanges of information across the organization – even when adding dotted lines between departments with formal reporting structures. Moreover, it does not reveal the dynamics of manufacturing, where different parts of the organization are working in temporary structures like project teams. The operations of a factory do no always match its official organization chart, as individuals, through the power of expertise or charisma, exert more influence than their position on the chart would suggest. Moreover, organization charts do not cover important power structures like worker councils and union representation. However, organization charts are useful analytical tools for understanding functions, relations, and department sizes. It is often one of the first things that are presented in factory tour introductions.

Figure 1.7 shows a typical kind of organization chart that could be found in a factory employing 1,000 people or more. As is common, it misses the key information of the number of employees in each box.

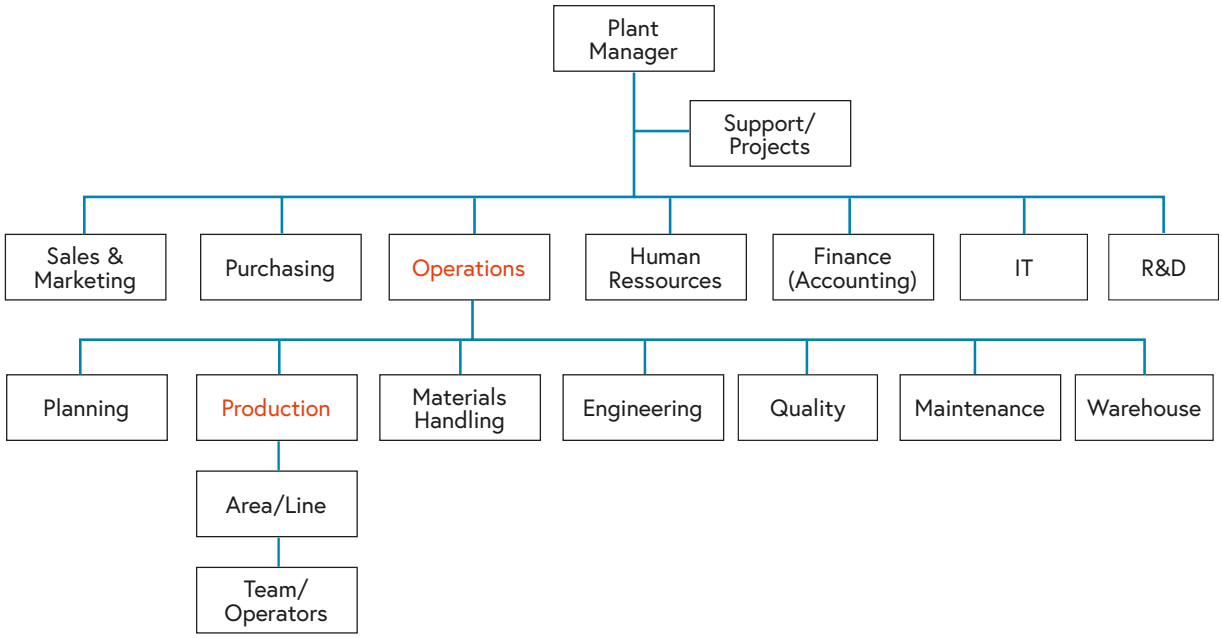


Figure 1.7 A typical organization chart in a large factory

The organization chart in Figure 1.7 includes a department called *Operations* and a series of other departments reporting directly to the Plant Manager. Depending on the tasks of the factory, not all these departments are always represented. For example, the *Sales & Marketing*, *R&D*, *Purchasing,* and *IT* departments can be located with the firm's headquarters in another city. *Human Resources* and *Accounting* or *Finance* can to a large extent be conducted outside of the factory organization, but they almost always need local representation and are therefore usually represented as departments, even if it is only one or a few people doing the task. Factories usually house *support departments* for administrative work, running investment and improvement projects, or following up Health, Safety, and Environment (HSE) regulations and tasks.

Some factories also have a separate *Warehouse* department that stores and manages finished goods and packs and ships products to customers. In smaller factories, this function can be co-located and co-managed with raw material inventories and covered by *Materials Handling,* sometimes under the name ofLogistics. In other factories again, the warehouse is located apart from the factory and not under the control of the factory management, but by a *Supply Chain* department in regional headquarters. The Supply Chain department also covers tasks such as external logistics and international shipping.

Stop and Think 1.4
Why is Planning often not organized under Production?

1.2.2. **The Operations department**

Exactly what is covered by the *Operations* department varies from factory to factory. In the example of Figure 1.7, the following functions are part of *Operations*:

- **Planning** prepares production plans based on forecasts and orders received from Sales.
- **Production** produces the parts or products according to the plans. It converts materials into goods.
- **Materials Handling** makes sure that Production has the materials it needs at all times. It is in charge of dock-to-dock in-plant logistics.
- **Engineering** is responsible for designing, building, and updating the factory's layout, machines, lines, and information systems.
- **Quality** is responsible for the methods used to ensure that products meet customer expectations.
- **Maintenance** is responsible for keeping machinery and tools in working condition and for repairs beyond what the operators can do themselves.
- **Warehouse** is responsible for storing and shipping finished products.



Figure 1.8 If Production is the race car driver, then Materials Handling is the pit crew

However a factory decides to organize, there will always be tensions among different interests. New plant and middle managers often come with new ideas for how to organize "better," so reorganizations are usual in factories. For example, in many plants, the *Materials Handling* function reports directly to the plant manager and is more remote from Production than in Figure 1.7. In these cases, the Materials Manager commonly pursues performance objectives for the department that may be counterproductive for the plant as a whole, like keeping material handlers busy. Instead, it should be making sure that *Production* has the materials it needs. This is the motivation to put it under the Operations Manager.

Within *Engineering*, there are roles that are definitely not part of Operations. Product development, for example, is often not even done on a manufacturing site. Production line design and new product introduction are usually done at the plant but not as part of Operations. Operations Managers, however, like to have control over engineers to tweak existing processes, improve task sequencing, vet improvement

ideas, or solve problems. If the engineers who do this work do not report to Operations, the Operations Manager needs to negotiate their availability every time he or she needs them, and commonly finds this structure cumbersome and unresponsive. The engineers, on the other hand, may find it frustrating to work in a line organization that does not understand the technical challenges they face, their need to keep up with new developments in the their field, and their career perspectives. This is the kind of dilemmas that matrix organizations attempt to solve by having engineers working for Operations report "dotted line" to a different Engineering Manager.

Box 1.2 **Centralized versus distributed support functions in Operations**

In almost every support function in Operations, the same tension exists between forms of organization. Should this particular form of support be provided by a department of specialists for the whole organization, or should it be distributed and put under the responsibility of line managers? You have the same dilemma in Quality, Engineering, Maintenance, Planning, IT, etc., and there is no single, universal answer.

The department of specialists is criticized as being bureaucratic, out of touch with the needs of the production floor, and managed in pursuit of goals that may not be in line with those of the business as a whole. But when you break it up and assign its members to production lines, you may make it impossible to assemble the critical mass of support needed to solve a particular problem, and you have these specialists report to bosses who do not understand their specialty, causing retention problems. There are also benefits of keeping specialists together: they can more easily have professional discussions and develop a community of experts. It is also difficult to balance the workload among distributed specialists. As a result, most factories end up having specialists organized in their own departments.

There are other ways than changing the organizational structures to ensure specialists are serving production. For example, locating the departments close to the shop floor and involving all departments in daily layered meeting structures.

1.2.3. The Production department

Within a factory, the Production department is special, as the only one that makes the products. Note that the greatest influence on a factory's performance is at the bottom of the chart in Figure 1.7, where operators and team leaders interact with first-line managers. That is where the products are made that customers buy.

First-line managers go by a variety of titles in different companies, which change over time. The "foreman" of the early 20th century had become the gender-neutral "Supervisor" by the 1980s, and then the less directive "Area Coordinator" or "Group Coordinator." The through-line of all these names is that it is the first rung of the management ladder – often populated by a mix of college graduates on their first job and senior operators promoted into this position.

The middle-level structure of Production is determined by the manufacturing processes and physical layout of the factory. If the process involves machining and assembly, with all the machine tools in one hall and all the assembly lines in another, then there will be a General Supervisor for Machining and one

for Assembly. If, on the other hand, they are organized by products, then there will be a General Supervisor per product or product family. And, if there are resources that cannot be dedicated by family, they usually go into a Shared Services department with its own manager – examples include expensive testing equipment, paint booths, or autoclaves. The point is to make clear interfaces with unambiguous areas of responsibility.

Box 1.3 **First-line manager to operator ratio in Toyota**

With between 40 and 100 operators to oversee, most plants' first-line managers have no time to do anything that is not directly related to their production numbers. While it is a widely held belief that having a low ratio of first-line managers to operators is "lean," it is not the way Toyota operates.



Figure 1.9 Operators finishing the four-millionth Yaris to roll of the production line in Toyota's factory in Onnaing, France, 2020 (Credit: S. Dhote / Toyota)

In Toyota car plants, these first-line managers have 4 to 6 teams of 4 to 6 operators each, or a minimum of 16 and a maximum of 36 operators, with an average that is just under 18. Operators are organized into teams of 4 to 6 people with a team leader.

The team leaders are operators with production tasks occupying 40% to 60% of the time available per unit at each operation, known as the *takt time*. In the rest of the time, they keep the pace, assist, or train the other operators, make sure materials and tools are available, keep records, coordinate changeovers, and call for help as needed. In addition to meeting production targets, first-line managers are expected to devote part of their time to leading improvement projects and organizing the professional development of their subordinates, which includes certifying their qualifications, helping them get additional training, and reviewing their progress twice a year.

Often, a factory can do better if the first-line managers have fewer operators reporting to them, for example 15 to 25, and their responsibilities would not be limited to meeting production targets. In addition, they would lead improvement projects and coach the professional development of their operators. They should be seen as part of management and be in daily, routine contact with production operators.

1.2.4. **Operators**

Although the operators populate the most important job in manufacturing, the gap between operators and management is marked in a variety of ways. For example, depending on industry and geographic location, it is not uncommon that different employee categories wear different work clothing. For example, shop floor employees can wear blue, the team leader perhaps light blue or gray, area management white, and technicians red. Higher-level managers and office workers usually wear their own

clothing. This choice is partly due to the need for different clothing in different jobs. In factories that require high levels of cleanliness, like the semiconductor industry, fine electronics, pharmaceutical industries or food processing, it is usual to wear white coveralls. Other factories are dusty or warm and require more robust protective coveralls. However, the use of *colors* to differentiate job categories is a strategic choice. In Japan and in many factories in the US, it is usually not possible to tell employees' positions from what they are wearing.



Figure 1.10 Shop floor workers clocking in and out

Another difference between shop floor operators and many office workers, is that it is common that operators are paid by the hour – clocking in and out – and receive overtime pay if asked to stay beyond their shifts. The higher-level managers and staff are "exempts," meaning that they receive a salary and do not measure their work time. Moreover, managers have many promotion opportunities in the company, while operators have few or none – unless they become managers.

It is important to recognize that not all operators are hired on the same type of contract. Some operators are permanent; others, temporary. Any manufacturing activity that is seasonal or cyclical has labor requirements that can vary by a factor of 3 or more over time. For example, fruit and vegetable processing is done during the harvest season, and project-based manufacturing is dependent on the scope and scale of projects. A rapid demand increase driven by market trends or external events – like a pandemic – may lead to the temporary need of more operators. Reliance on temporary staff, however, not only compounds the difficulty of hiring, but prevents skills growth in the workforce. Demand seasonality is an external cause. The problem is not created by factory management but it can cripple performance improvement. It will not be solved unless management pursues a different approach to human resource management, aimed at developing a smaller, permanent force of skilled workers, focused on meeting the demand during the high season and on improvements/skills development during the low season.

When we realize that operators, not managers, make what customers buy, we should find a way to create good operator jobs. A sustainable approach is to take full advantage of the characteristics of humans that make them special as production resources:

- The ability to learn.
- The ability to respond to unforeseen events.
- The intelligence to solve problems.
- The creativity to find new uses for existing devices and invent new ones.
- The ability to care for other humans and help create a good work environment.
- Physical strength, dexterity, and endurance.
- Ethics that enable them to act in the best interest of a group.

Releasing these human abilities fully, each employee has to act as if the survival of the company depends on his or her individual behavior. There is no sure recipe to make employees feel this way but there are many ways to make sure they don't. Treating operators with disrespect, making them work in unsafe environments, or not paying a fair salary for their work effectively hinder the operators' human abilities to be put to good use.

In many factories, work is designed by specialists and time studies are used to set standards intended to keep operators from colluding to slow down. Such factories can be both effective and efficient, but they will never turn into centers of excellence, nor are they particularly good places to work. There is, however, another tradition in industrial engineering. Companies like Volvo in Sweden and Toyota in Japan study work design with the purpose of improving it, in collaboration with those who do it.

Teruyuki Minoura, previous CEO of Toyota Motor Manufacturing North America, Inc., quoted Eiji Toyoda elaborating on Benjamin Franklin's point that "time is the stuff life is made of,"[4] in an address to the World Class Manufacturing Forum in 2002:

> *Mr. Eiji Toyoda tells the following story of a Buddhist monk: "The monk said, 'The human lifetime is made up of an accumulation of time. So one hour is a part of that person's lifetime; in other words, it is his or her life itself.' So, when we hire our employees to work for our company, we purchase their life for the duration of the time they spend with us. The employees are offering a very important part of their life to us. If we don't use their time effectively, we are wasting their lives."*

Taking this perspective is not philanthropy. Specialists can improve *one* operation out of hundreds performed in a factory. Attending to the required level of detail on *all* operations is beyond the capacity of any reasonably sized engineering department. The design of operator jobs for the whole factory can only be done by engaging the line organization, from the operators who do the jobs to their first-line managers.



Figure 1.11 Operators are the most important people in manufacturing

4 Franklin, B., Franklin, W. T., & Duane, W. (1834). *Memoirs of Benjamin Franklin*. M'Carty & Davis.

Box 1.4 **Building T-type specialists in support roles through job-rotation**

Some companies intentionally develop T-type specialists and leaders. This means having a deep expertise in one functional area (e.g., Engineering, Purchasing, or Sales), and a broad expertise in at least two other functional areas.

The people in support roles should render services production needs for today or tomorrow. Most companies staff departments with specialists who have not necessarily worked in production. It is in fact common for people to move from one support department in one company to the same department in another (e.g., Engineering, Accounting, or Logistics), allowing individuals to spend their entire careers in the same support function.

While this is expected to result in a higher level of specialized expertise, it does not promote mutual understanding and cooperation across departments. Among the many possible countermeasures is the systematic rotation of some employees between departments every few years. It is common in Japan. The person you encounter in Production Control may have been a Design Engineer for 5 years and, as a result, understands engineering issues. Working in the Quality department does not have to be a lifetime commitment. You can come from Production, spend a few years in Quality, and go on to Maintenance. People who rotate in this fashion may have less depth in specialized knowledge but more on how the organization works. Incidentally, this practice makes them both more valuable to their current employer and less to others.

Of course, nothing works perfectly and everywhere. Rotating high-potential individuals in this fashion is done with varying degrees of effectiveness worldwide. In some companies, the "tours of duty" are too short for the individuals to be committed, and they are rewarded for not making waves. In such cases, the rotation policy may perpetuate dysfunction.

In larger organizations, the policy of rotating some of the professionals between departments is one approach. A matrix structure is another. In it, you have each professional reporting "solid-line" to an operational manager and "dotted-line" to a functional manager of the same specialty, who can provide guidance in career development and skills maintenance. For example, an engineer supporting a production line reports to the line manager, but also stays connected with an engineering manager to stay current with technology, attend conferences, receive continuing education, and plan a career.

On a case-by-case basis, you have to consider the nature of the support needed. How deep is the required specialized knowledge? You would not rotate people between open-heart surgery, rocket science, and baseball, but maybe you can between Production Control and Quality. These departments serve each other and communicate with each other, so understanding the other side is invaluable.

1.3. The history of manufacturing

The history of manufacturing teaches us lessons that are applicable today. It deserves many volumes on its own, and the point here is only to show some milestone examples as a starting point to dig further.

The following development phases can be separated in the history of manufacturing:

1. Pre-industrial production
2. Craft production
3. The Industrial Revolution
4. The American System of Manufacture
5. Mass production
6. The Toyota Production System and lean manufacturing

In addition, the evolution and diffusion of Information Technology (IT) and Operations Technology (OT) since the invention of the computer in the 1940s has opened up many opportunities. While some argue that this adds up to a "4th Industrial Revolution," also known as "Industry 4.0," it is not a consensus view. These phases have been enabled by progress in technology and/or concepts for organizing manufacturing.

Note that the advent of a new phase does not *replace* previous phases but add to them. Instead, they coexist, often in an interconnected web of manufacturing activities. The history of manufacturing is not as well known as the history of governments, wars, or inventions. We know the history of technology better than the history of manufacturing. Regarding, for example, microscopes, schoolchildren may hear about the 17th century Dutchman, Antonie van Leeuwenhoek, who made his own microscopes, and whose life is well documented. But who, in the aftermath, set up factories to make microscopes for others to use? And how did they organize production? This is not widely known and is not easy to find. When studying history you need to approach all sources of information with a critical mind.

1.3.1. Production in the pre-industrial area

We know very little about how things were manufactured before the mid-1700s. Archeology tells us that the oldest objects made by working copper date back to 6500 BCE in Çatalhöyük in what is today Turkey, and that the extraction of metal from ore was first done in 5000 BCE but it doesn't tell us how. In *Poiesis*,[5] Peter Acton tells us about manufacturing in Athens in the 5th century BCE. We know that the Vikings in Scandinavia manufactured wooden ships around 1000 CE that could travel more than a thousand nautical miles in rough seas, but we do not know how the manufacturing processes were organized. The Venetian Arsenal outside Venice in Italy employed about 16,000 people in the mid-1500s, who together were able to produce nearly one ship a day in its peak years. However, we do not know exactly how they did it. A few noteworthy books on metallurgy provide detailed insight into mining and manufacturing processes of metals in the pre-industrial era. Three examples are *Ein nützlich Bergbüchlein*[6] (The useful little mining book) in 1500, *De la Pirotechnia*[7] published in Italian in 1540, and *De re metallica*.[8]

5 Acton, P. H. & Acton, P. (2014). *Poiesis: Manufacturing in Classical Athens.* Oxford University Press.

6 Rülein von Calw, U. (1955). *Ulrich Rülein von Calw und sein Bergbüchlein* (Pieper, W., Ed.). Akademie-Verlag. (Orig. work published 1505)

7 Biringuccio, V. (1990). *The pirotechnia of Vannoccio Biringuccio.* Dover Publications.

8 Agricola, G. (2013). *De re metallica.* Dover Publications.

Between 1751 and 1772, the *Grande Encyclopédie*[9] came out in France, with more than 2,000 drawings showing the manufacturing processes used at the time, Figure 1.12 shows an example..

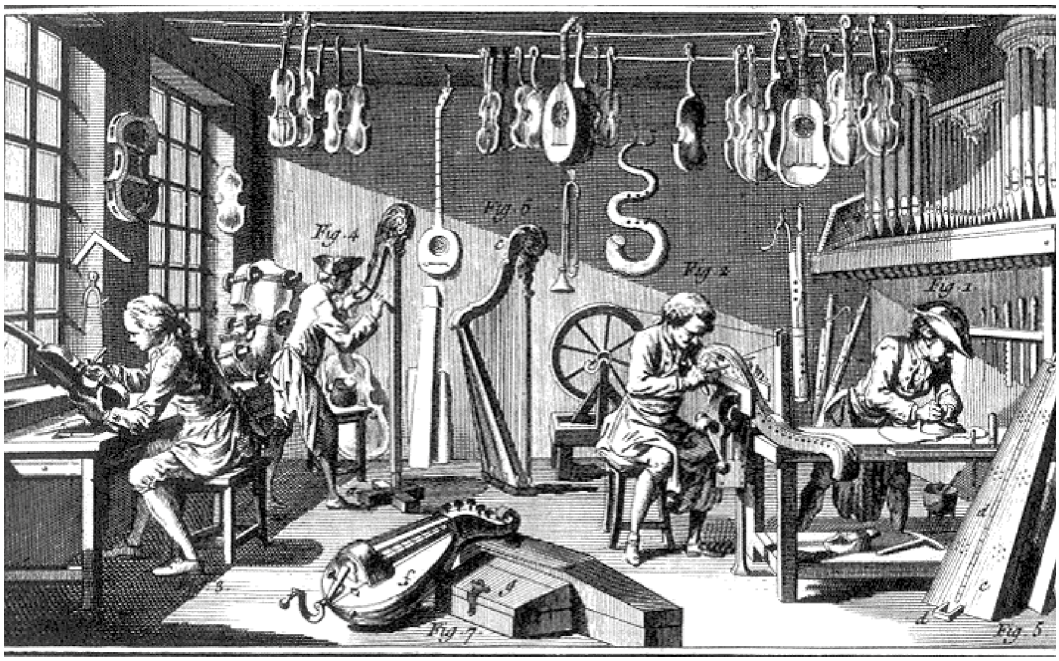


Figure 1.12 An illustration of manufacturing of musical instruments in the *Grande Encyclopédie*

Box 1.5 **The ceramics of Jingdezhen and early industrialization in China**



Figure 1.13 China has produced porcelain for millennia

Jingdezhen is a region of China that has been known as a center of fine ceramics manufacturing for 1700 years, which raises the question of how such a feat could be possible through four imperial dynasties and a 20th century of revolutions and wars. From the mid-14th century, Jingdezhen become one of the world's first industrial towns when beginning volume production of underglaze blue porcelain, as shown in Figure 1.13.

One possible explanation proposed for this extraordinary longevity is an industry structure in which each step of the process, from clay extraction to firing, has been performed by different businesses, and that these businesses were able to expand and contract independently, where a vertically integrated organization would most likely not have survived.

9 Diderot, D. (1993). *A Diderot pictorial encyclopedia of trades and industry*. Dover Publications.

1.3.2. **Craft production**

We know that the dominant mode of manufacturing before the Industrial Revolution was craft production. Craft production is the process of manufacturing by hand, usually with the help of simple tools. The craftspersons usually works in small workshops. Examples are a black smith forging swords, a jeweler making jewelry, a wood smith making paddle boats, and a tailor making clothes.

The Middle Ages in Europe gave us the guild system, which provided for the training of apprentices in a variety of trades. Apprentices or companions traveled between cities to learn traits from various masters. They would then make and defend their "masterpieces" before a jury of masters in order to become masters themselves. This system, however, had a dark side: It restricted the diffusion of know-how and forbade innovation in order to protect the positions of established master craftsmen. This system has not completely disappeared in the 21st century. Guilds still exists in Germany and France, operating schools and granting degrees. The guild mentality pervades other professions where know-how cannot be protected by patents, like sushi chefs or martial arts instructors.

Although its dominant position in the manufacturing sector is long gone, craft production is still alive. Crafted products are sought-after for their uniqueness and character and sometimes for their quality. To build a successful business on craft production, you need a sort of talent or skill that sets you apart from the competition. This skill is what makes you a *craftsperson*, or artisan. See Figure 1.14 for four examples of modern craft production.



Figure 1.14 Modern craft production: a carpenter, a potter, a glass blower, and a fashion designer.

Craft production is characterized by the following:

- Skilled, often guild-certified artisans carrying out all critical production steps.
- Skill often learned through years of apprenticeship with a guild master.
- Artisans are usually in direct contact with customers, employers, if any, and suppliers.
- No economies of scale and very low production volumes.

- No, or few, industry standards.
- Use of general-purpose tools.

1.3.3. The Industrial Revolution

The Industrial Revolution started in the textile and iron industry in the Northern parts of England in the latter half of the 18th century. This history is well documented in many articles and books, and is taught at the primary-school level in many countries. The most important developments during the Industrial Revolution should be common knowledge for factory engineers and managers. The illustration shown in Figure 1.15, published by Samuel Griffiths in 1873, gives an impression of the early days of industrialization in England.

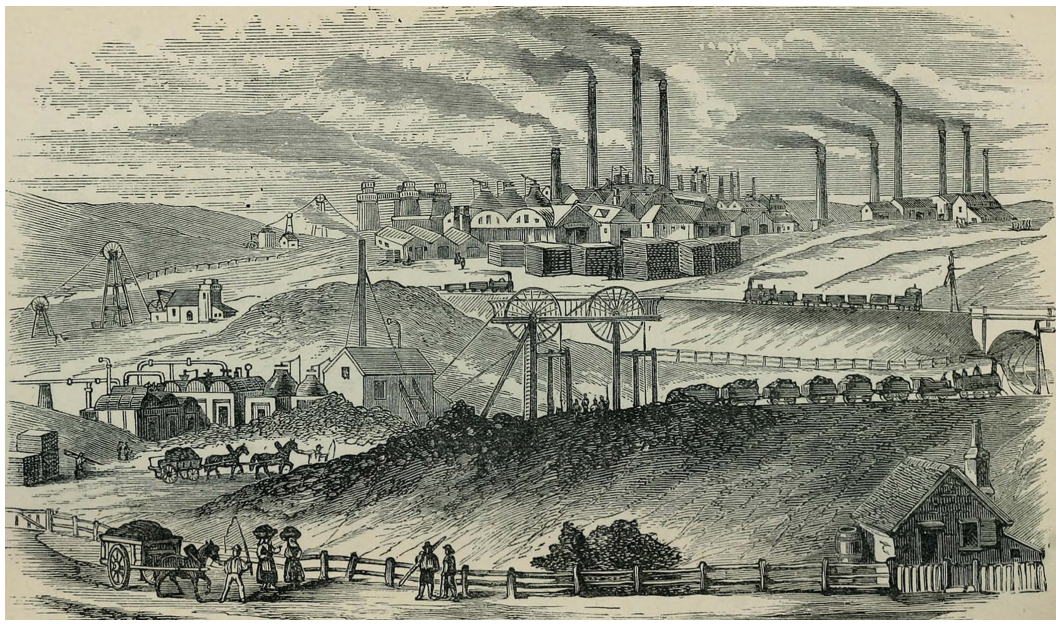


Figure 1.15 The Industrial Revolution in the Black Country west of Birmingham, England

James Hargreaves' development of the Spinning Jenny in 1764 has become a symbol for the start of the Industrial Revolution (Figure 1.16). The further development of the spinning machine and related technologies by many others was central in the establishment of some of the world's first textile factories using mechanized spinning in the 1780s. In 1768, Richard Arkwright patented the first water-powered spinning frame (Figure 1.16). The water frame was installed in Arkwright's cotton mill at Cromford in Derbyshire, which was one of the first factories purposely constructed for machinery rather than manual work. Arkwright also introduced a new manufacturing management practices that is still valid: People were employed instead of contracted and the working day was determined by the clock rather than daylight hours.

Note that concurrent developments in the systems of transportation, communication technology, and banking were also central factors that enabled industrialization to take place. For example, James Watt perfected the steam engine and made use of one in a cotton mill in 1785, and, by that, the replacement of human, animal, and water power was complete.

Coal was cheap and factories could now be placed almost anywhere. Unlike water power, steam power was also available year round, which made it possible for work in factories to become a permanent occupation, as opposed to a seasonal activity supplementing farming.

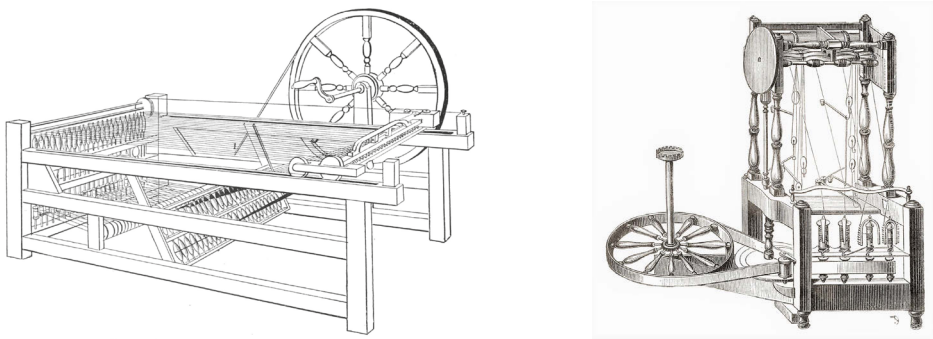


Figure 1.16 Mechanized spinning technology from the 1760s: Spinning Jenny and Water frame.

Mechanized textile production spread to continental Europe and the United States in the early 1800s. During this period, the predominantly agrarian and rural societies in Europe and America developed into industrial and urban societies

Industrialization led to a shift away from simple hand tools in small workshops to powered and special-purpose machinery in factories. The introduction of factories made it easier to divide work tasks between employees. Whereas craft production is characterized by the craftsperson carrying out all the critical production processes, a division of labor among several workers characterizes industrial production. One advantage of this system is that cheap, unskilled workers could be employed to carry out specialized and repetitive tasks.

The introduction of factories with the *division of labor principle* radically improved productivity (see Box 1.6 for the famous pin factory example of economist Adam Smith) but not the quality of life of the production workforce. They were subject to brutal discipline and lost the connection with the end products, which had been a characteristic of craft production. Karl Marx later called this "alienation."

Industrialized production is characterized by the following:

- Production takes place in factories.
- Separation of sales and manufacturing.
- Use of machines with external power sources.
- Factory owners employ a mix of skilled and unskilled labor.
- Division of labor.
- Economies of scale are achieved with high production volumes.

Stop and Think 1.5

With the objective to stop the Industrial Revolution in England, the "Luddites" used violence against factories and machinery. Two hundred years later, an American mathematician named Ted Kaczynski became notorious as the "Unabomber" for sending bombs by mail to protest against "industrial society." Why do you think someone would want to stop technological progress?

Box 1.6 **Division of labor: The pin factory example of Adam Smith**

To learn about the magnitude of improvement made possible by the Industrial Revolution, consider the below excerpt of the famous pin factory example from the first chapter of Adam Smith's book Wealth of Nations (1776).[10]

> *"To take an example, therefore, from a very trifling manufacture; but one in which the division of labor has been very often taken notice of, the trade of the pin-maker; a workman not educated to this business (which the division of labor has rendered a distinct trade), nor acquainted with the use of the machinery employed in it (to the invention of which the same division of labor has probably given occasion), could scarce, perhaps, with his utmost industry, make one pin in a day, and certainly could not make twenty. But in the way in which this business is now carried on, not only the whole work is a peculiar trade, but it is divided into a number of branches, of which the greater part are likewise peculiar trades.*
>
> *One man draws out the wire, another straights it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving, the head; to make the head requires two or three distinct operations; to put it on is a peculiar business, to whiten the pins is another; it is even a trade by itself to put them into the paper; and the important business of making a pin is, in this manner, divided into about eighteen distinct operations, which, in some manufactories, are all performed by distinct hands, though in others the same man will sometimes perform two or three of them. I have seen a small manufactory of this kind where ten men only were employed, and where some of them consequently performed two or three distinct operations.*
>
> *But though they were very poor, and therefore but indifferently accommodated with the necessary machinery, they could, when they exerted themselves, make among them about twelve pounds of pins in a day. There are in a pound upwards of four thousand pins of a middling size. Those ten persons, therefore, could make among them upwards of forty-eight thousand pins in a day. Each person, therefore, making a tenth part of forty-eight thousand pins, might be considered as making four thousand eight hundred pins in a day. But if they had all wrought separately and independently, and without any of them having been educated to this peculiar business, they certainly could not each of them have made twenty, perhaps not one pin in a day; that is, certainly, not the two hundred and fortieth, perhaps not the four thousand eight hundredth part of what they are at present capable of performing, in consequence of a proper division and combination of their different operations."*

10 Smith, A. (1904). *The wealth of nations* (Cannan, E., Ed.). Methuen & Co. (Original work published in 1776)

1.3.4. The American System of Manufacture

For the first hundred years of the Industrial Revolution, parts in an assembled product were usually not interchangeable. In the Smithsonian museum in Washington D.C., visitors can see Singer sewing machines built until the 1870s with the serial number of the end-product engraved on every part, indicating that the parts were not expected to fit on any other machine. These numbers do not appear on later machines.

The early factories had neither gauges nor standard processes and did not have the concept of critical dimension or tolerance. As a result, dimensions varied so much from part to part that they were not interchangeable. This problem was most apparent for dysfunctional weapons on the battlefield.

Already in 1765, the French general de Gribeauval envisioned a "uniformity system" where firearms could be repaired in the field with perfectly interchangeable spare parts. Gribeauval's work was focused on cannons, and he is now primarily remembered as the designer of the artillery system used 40 years later in the Napoleonic wars. A decade later, gunsmith Honoré Blanc took up the challenge of making muskets with interchangeable parts. In 1785, he demonstrated his results to American ambassador Thomas Jefferson. Later, as US president, Jefferson launched the development effort that was eventually successful.

In the 2020s, sensors can capture just about every parameter imaginable. In the late 1700s, on the other hand, thermometers to measure the appropriate temperature for forging iron had yet to be invented, and instructions called for the metal to be "cherry red." In such a context, interchangeable parts were an almost insurmountable technical challenge. In 1807, the French government terminated this program for social rather than technical reasons. The gunsmiths of Saint-Etienne, France, had learned to forge a musket barrel from a slab of iron with hand tools. To preserve their positions, they organized to prevent the implementation of Blanc's system of proto-machine tools. While not formally a medieval guild, they opposed innovation as the guilds had and were eventually successful in stopping further developments in France.

Interchangeable parts technology *was* developed in the US, and returned to France half a century later as the "American System of Manufacture." It was enabled by engineering drawings with critical dimensions, tolerances and allowances, the invention of machine tools, fixtures and jigs, and overall progress in basic metallurgy, metalworking, and metrology.

This development took place from 1800 to the mid-1850s, primarily in government armories at Springfield, MA, and Harpers Ferry, VA. The American weapons industry then implemented it. For example, Colt's revolver factory produced over 200,000 Model 1860 revolvers from 1860 to 1873 following the American System of Manufacture, and it gradually found its way into other industries. Figure 1.17 shows a metallic cartridge rifle manufactured by Springfield Armory in Boston, MA. The metal parts are interchangeable. The rifle shown has no serial number. The armory produced 341 units of this model.

Interchangeable parts technology is so thoroughly assimilated in manufacturing today that it is no longer a topic in its own right and that it is difficult to understand why it ever was controversial and disruptive. Without this foundation, however, developments like mass production or the Toyota Production System would not have been possible. In 1908, Sakichi Toyoda hired American professor Charles A. Francis to teach the methods in his loom factory.



Figure 1.17 The Springfield Model 1870 used interchangeable parts so they could be repaired

In the 20th century, the focus shifted to process capability – that is, the ability of processes to hold the required tolerances reliably. It led to the development of Statistical Process Control (SPC) in the 1920s, the application of statistical Design of Experiments (DoE) in manufacturing in the 1950s, and other methods since.

1.3.5. **Mass production**

"Mass production" is a generic term coined at Ford for the company's revolutionary production system developed for the Model T (Figure 1.18). Ford's Samuel Crowther provided the following description in the *Encyclopedia Britannica Supplement* of 1926:

> *"Mass production is not merely quantity production, for this may be had with none of the requisites of mass production. Nor is it merely machine production, which also may exist without any resemblance to mass production. Mass production is the focusing on a manufacturing project of the principles of power, accuracy, economy, system, continuity, and speed."*



Figure 1.18 Ford Model T

The main point is that mass production is a systems approach. The key innovation at Ford was the assembly line. Ford's assembly line may not have been the first in history, but it was the first in the car industry and became the most influential worldwide in all manufacturing sectors.

According to historians, Ford engineer William 'Pa' Klann picked up the idea of the assembly line during his visit to a slaughterhouse in Chicago. Since 1867, it had been a usual practice to use hanging conveyors as a "disassembly line" for cutting carcasses in the Chicago meat packing industry.

Henry Ford wanted to manufacture a car cheap enough for his own factory workers to buy, and moving the parts rather than the human worker seemed to be an effective idea. The assembly line brought the car to the workers instead of workers moving around to the cars. The moving assembly line was introduced in the Highland Park Ford Plant in 1913. It cut the throughput time from 728 minutes to 93 minutes, which was less than the paint of that era needed to dry. By continuously lowering production costs, while maintaining quality, Ford created the market for his cars. Figure 1.18 shows the Ford Model T, also known as T-Ford and Tin Lizzie. It was the first car produced on assembly lines. Ford continued its production until 1927 by which time 15 million Model Ts had been made. This record, for a single model, was later beaten by the Volkswagen Beetle, with 21 million units made between 1950 and 2003.

Mass production was not without problems. It was repeatedly under attack from workers and observers for providing repetitive, hard, and dull jobs on the assembly line. In *Modern Times* (1936), Charlie Chaplin criticized the working conditions on many assembly lines. During this time, the understanding of human work increased substantially.

In addition to the characteristics of industrial production and the American System of Manufacture mass production was characterized by the following:

- A focus on flow, through the moving assembly line.
- Use of time and motion studies.
- Limited training needed for repetitive and standardized assembly jobs.
- Make-to-stock production serving a rapidly growing market.
- Dedication to a single product, the Model T, a competitive advantage in the 1910s but a handicap in the 1920s when the Model T became obsolete.
- Vertical integration. The River Rouge plant turned ores, sand, and other materials mined from the earth into finished cars. The plant still exists but all the processes upstream from final assembly are now owned and operated by suppliers. The Volkswagen main factory in Wolfsburg, Germany, was inspired by the Rouge.
- One-man rule by Henry Ford. A sustainable management structure was not put in place until his grandson, Henry Ford II, took over after World War II and brought in a group of executives who came to be known as "the Whiz Kids." Best known among the Whiz Kids is Robert McNamara, Ford President in 1960, U.S. Secretary of Defense for Presidents Kennedy and Johnson, and President of the World Bank.

As implemented, mass production decoupled assembly from sub-assembly production through inventory. High-volume, dedicated lines are still used today for products with sufficient demand. Unlike mass production lines, they are synchronized with feeder lines and the supply chain to avoid the accumulation of work-in-process (WIP).

Alfred P. Sloan,[11] who led General Motors to surpass Ford and any other company in the world by the 1930s, added another idea to the mass production system. Sloan differentiated GM from Ford by offering a broad selection of cars tailored for the changing tastes and fortunes of consumers and by adding the latest technology and annual styling changes to the car models. The five GM models at that time were Chevrolet, Pontiac, Oldsmobile, Buick, and Cadillac, and they represented a "ladder of success" that families would climb as their financial situation improved. The yearly model changes "brought the laws of Paris dressmakers to the car business." GM invented "planned obsolescence" and was able to keep customers for a lifetime.

Sloan also set up a decentralized, meritocratic management structure that, for decades, was studied as a model to follow as, for example, in Peter Drucker's *Concept of the Corporation*.[12] By the 1980s, however, this system had lost its vitality and effectiveness.

11 Sloan, A. (2015). *My years with General Motors.* Currency.

12 Drucker, P. (1993). *Concept of the corporation*. Transaction Publishers.

Box 1.7 **Frederick W. Taylor and Scientific Management**

Frederick Winslow Taylor's work on "Scientific Management" around the turn of the 20th century is inevitable material in any management textbook. His work predates the Ford assembly line, and its designers do not acknowledge his influence. His theories are based in part on the work he did for Bethlehem Steel Works in Pennsylvania (Figure 1.19).



Figure 1.19 Bethlehem steel works, Bethlehem, Pennsylvania, 1881 (Credit: Watercolor by Joseph Pennell)

His approach posits that work processes must be studied and measured systematically to prevent workers from colluding to curtail output. Taylor – the world's first consultant – and his associates studied how work was done, and how it affected worker productivity. Taylor observed that the work methods used in factories were effectively controlled by production workers, and undertook to put an end to this by what he called "Scientific Management."

Scientific Management divides work between the "heads" and "hands" of production; the heads (management) should plan and control the production carried out by the hands (operators).

Scientific Management rests upon the following four core principles (Taylor, 1911, p. 27).[13]

- Develop a science for each element of a worker's job, which replaces the old rule-of-thumb method.
- Scientifically select and then train, teach, and develop the worker, whereas in the past he or she chose how to do the work and trained as best as possible.
- Heartily cooperate with workers to ensure that all the work is done in accordance with the principles of the science that has been developed.
- Divide work and responsibility equally between the management and the worker. Managers take over all work for which they are better fitted than the workers, while in the past almost all of the work and the greater part of the responsibility were thrown upon the workers.

Taylor is mostly known for time studies, but he also invented high-speed machining and categorized the functions that are now organized as support departments in manufacturing firms.

13 Taylor, F. W. (1911). *The principles of Scientific Management*. Harper & Brothers.

Box 1.8 **Frank and Lillian Gilbreth**

While less known than Frederick Taylor, Frank and Lillian Gilbreth have actually had a more positive and long-lasting influence over job design. Instead of attempting to police and pressure workers to work harder, Frank and Lillian Gilbreth were focused on making their jobs easier.

They pioneered the use of motion pictures to analyze operations, designed the way hospital operating rooms have been working for the past 100 years, and invented predetermined time standards. Lillian Gilbreth, on her own after Frank died in 1924, designed the modern American kitchen (Figure 1.20) and invented industrial psychology. She was the only pioneer of Industrial Engineering with a PhD, which she earned in applied psychology from Brown University in 1915.



Figure 1.20 Lillian Gilbreth standardized work heights and appliance locations

1.3.6. **The Toyota Production System and lean manufacturing**

The Toyota Production System (TPS) first became known to the outside world when Toyota started training suppliers in it in the late 1970s. Since, by definition, it is the way Toyota produces, it is still a work in progress, and learning it is a never-ending process as the content keeps changing. As described by Takahiro Fujimoto,[14] it was not the result of a grand design but instead emerged over time as Toyota people devised countermeasures and solutions to a succession of crises.

Toyota's rise from a startup near bankruptcy in 1950 to a world leader making almost 10 million cars per year is largely credited to TPS. This raises the question of what specifically makes this system so

14 Fujimoto, T. (1999). *The evolution of a manufacturing system at Toyota*. Oxford University Press.

different from mass production? TPS is covered in detail later and various elements of it appear throughout the book. Here it is sufficient to summarize a short list of key TPS strategies:

- JUST-IN-TIME. Delivering the right items in the right quantities at the right time without relying on inventory to do it.
- JIDOKA. Making production machinery progressively easier for operators to use through incremental automation and making it stop automatically when it malfunctions.
- GEMBA. Focusing on the actual place where the work is done. In a factory, it is the production shop floor.
- FLOW. Removing interruptions in the movements of materials and information from the supplier, through the plant, and to the customer.
- KAIZEN. Working every day to make operations better.

JUST-IN-TIME. Just-in-time (JIT) implies short lead times from suppliers to the factory, between operations inside the factory, and between the factory and customers. It is as much about quality as about inventory holding costs. With JIT, defects are promptly discovered, corrected, and reported to the source to prevent recurrence, in short cycles of learning.

JIDOKA. Jidoka, written 自動化, means "automation" in everyday Japanese, but Toyota writes the second kanji differently, as 自働化, adding the radical 人, for human. As it does not change the pronunciation, it is an untranslatable pun but Toyota's jidoka is often translated as "automation with a human touch."

While the original approach to manufacturing automation is machine-centered with people as an afterthought, the Toyota approach is human-centered and incremental. Regardless on the amounts invested in machinery, its effectiveness hinges on the ability of people to use it. In the English-language literature, Toyota's jidoka is often reduced to one of its features: Machines' ability to stop automatically when they malfunction. It does not do justice to this much richer concept.

GEMBA. Gemba is a Japanese word meaning "the actual place." In manufacturing, it is the factory floor. Market prices for products and the costs of materials, energy, and outside services are largely outside the control of manufacturers. The TPS approach is to pursue profitability by improving work on the production shop floor, as opposed to raising prices, pressuring suppliers for concessions, or buying new technology.

FLOW. Materials should ideally flow along paths like water in a river, at a steady pace, with each unit at each operation followed by another. In a river, the flow is broken by dams and locks; in a factory and its supply chain, by batching for processing or transportation and by all sorts of disruptions.

In production, flow is accomplished by, wherever possible, using machines that process one piece at a time, laying out machines and workstations so that the output of one feeds directly into the next one, and locating operator interfaces so close to one another that an operator can easily attend to a *sequence* of machines or workstations. Between production lines inside the factory and in the supply chain outside of it, flow can be supported by pull signals that prevent the accumulation of unnecessary stocks.

This not only reduces lead times, but also accelerates the detection of quality problems. Note that factories reap more benefits from improving what they do all the time than what they do rarely and randomly. Rather than having a one-size-fits-all approach to making products, regardless of the demand

structure, different policies for different volume categories should be applied. The same logic applies to in-plant logistics, supply chain logistics, or maintenance.

IMPROVEMENT (KAIZEN). Kaizen (改善) is a Japanese word meaning "change for the better" – that is, improvement. In the lean literature it is usually called *continuous* improvement, indicating that it is incremental and must be constantly pursued, with no end.

Mass producers believe their factories to be, if not optimal, at least *Pareto optimal*, meaning that performance in one dimension could only be improved by making others worse. For example, they see trade-offs in quality: It could be enhanced, but by increasing costs and delaying deliveries. A key postulate of TPS is that, on the contrary, there is no such trade-off, and that you can always pursue simultaneous improvement in all dimensions of performance, which is what eliminating waste actually means.

Note that *improvement* and *optimization* are different mindsets. Once you have optimized a process it can no longer be improved. On the other hand, once you have improved it, the next step is to improve it again. The improvement mindset is that there is no such thing as an optimal factory. All real factories can be improved, and improving them is an integral part of their employees' jobs.

It is a matter of consensus that TPS is still the best way to manufacture cars. Also in many other assembly processes, TPS remains the best-known way to organize manufacturing operations. It has been successfully adapted in industries ranging from commercial aircraft to furniture, medical devices and surgical robots. Yet few other companies proudly proclaim that they are applying TPS.

There are several reasons for this:

- Car manufacturers do not want to advertise that they are using a competitor's system.
- Outside the car industry, it is not obvious that TPS is the best way. By experience, managers and engineers can be quick to dismiss TPS if they are making frozen lasagna, airplanes, or any other products but cars.
- Toyota does not have a monopoly on good ideas. Quick Response Quality Control (QRQC), for example, is an approach to quality management from Nissan that has been widely adopted but it would be inaccurate to describe it as part of TPS.

As consultants helping companies implement the TPS approach have multiplied since 1980, their need for branding has led to a profusion of different names, including "World Class Manufacturing," "Synchronous Manufacturing," "Operational Excellence," and, since 1990, a variety of expressions starting with "lean." The term "lean production" was introduced by John Krafcik in 1988[15] and popularized through the book *The Machine that Changed the World* by James Womack, Daniel Jones, and Daniel Roos in 1990.[16] It then became "lean manufacturing," "lean thinking," the "lean enterprise," "lean management," "lean six sigma," "lean 4.0," or just "lean."

While Krafcik used "lean production" as a generic label for TPS, the approaches sold under the various lean labels diverged over time from TPS. Many "lean implementations" today are centered on tools that are unknown at Toyota, like kaizen events, have been rejected by Toyota, like six sigma, or play a minor

15 Krafcik, J. F. (1988). Triumph of the lean production system. *Sloan Management Review, 30*(1), 41-51.

16 Womack, J. P., Jones, D. T., & Roos, D. (1990). *The machine that changed the world*. Rawson Associates.

role in TPS, like Value Stream Mapping. All these methods and tools can at best bring some local benefits, but not the systematic change and continuous improvement culture that Toyota has become known for.

Rather than using generic labels, many global companies have chosen to follow Toyota's lead in a different way, by naming and building their own company-specific production systems. Examples are the Bosch Production System, the Nissan Production Way (NPW), the Volvo Production System (VPS), or the AGCO Production System (APS).

1.3.7. Information Technology and Operations Technology

Today, no manufacturer in the industrialized world operates without computers. But computers didn't exist as commercial products until the 1950s. Until then, business transactions and bookkeeping were done manually with pen and paper, aided with calculators or perhaps the typewriter since the late-1800s. Production planning, scheduling, and control was done on paper spreadsheets, or followed simple rules like reorder point. Engineering drawings were handmade and engineering calculations done with slide rules and books of tables by humans whose job title was "computer."

The computer was invented in the late 1940s and has been the most consequential innovation since the printing press. It has changed the way we earn and pay money, play games, read books, get daily news, drive cars, write, take pictures, make and watch movies, design products, etc., and, along the way, it has also found its way into manufacturing, as exemplified in Figure 1.21.



Figure 1.21 Computers used to control machines

Why is the computer so important for manufacturing? Because, as Information Technology (IT), it can do tasks related to production planning, scheduling, and control that people are not able to do, not willing to do, or could not get right. The use of computers enable management to handle a much larger complexity than they could with manual processes. As Operations Technology (OT), computers can automate and track physical work.

In particular, IT supports manufacturing with the following tasks:

1. Plan and schedule production.
2. Maintain master data like bills of materials, work instructions, process programs, or the roster of employees.
3. Keep historical records for analysis and traceability. In the domains that are closest to production, computers have been a great success. For example, Computer-Aided Design (CAD) for products, programmable controllers for machines, or Warehouse Management Systems (WMS).

OT refers to functions that are closer to production equipment and processes:

1. Execute production, including automation of tasks.
2. Track status of production.

Success has been more elusive in the high-level management functions that have seen a succession of generations of systems since the 1960s: Materials Requirements Planning (MRP), closed-loop MRP, Manufacturing Resource Planning (MRPII), Enterprise Resource Planning (ERP), and cloud-based ERP. They have grown into a multi-billion-dollar industry. Despite all the available IT, a majority of manufacturers still rely on Excel spreadsheets to work around the limitations of their IT systems.

People often think of computers as machines like any other but the reality is that they are qualitatively different, and that programmability allows them to outperform dedicated tools. Their hardware configurations make them smartphones, game systems, laptops, or industrial controllers, and only the imagination and talent of programmers limit the range of services they can render well. What makes the computer different is that it is a *universal machine*.

Box 1.9 General Motors' lights-out strategy in the 1980s

A radical alternative to the Toyota Production System suggested in the 1980s was computer-integrated manufacturing (CIM), which promised to use computers to control the entire production process. At that time, Roger Smith was the Chairman and CEO of General Motors Corporation (tenure from 1981 to 1990). When he left, he received the distinction of "Worst American CEO of All Time" from CNBC. Why?

Facing competition from foreign automakers such as Toyota, one of the prominent strategic decisions of Smith was to modernize GM using advanced technology. The vision was "lights-out" factories where the only employees were those supervising the robots and computers. Through the 1980s, GM spent more than $90 billion to reinvent itself. In its search for automation robots that could be installed in the factories, GM made a joint venture with the Japanese robot manufacturer Fujitsu-Fanuc. With that, GM became the largest industrial robot manufacturer in the world. The vision did not materialize. There are reports of robots welding doors shut and robots painting each other instead of the cars.

During Roger B. Smith's tenure, GM market share in the US went from 46% to 35%. The factories in Detroit went from among the lowest cost factories to the highest cost factories. The amount of money spent on automation would have been sufficient to buy several of the toughest competitors such as Toyota, Nissan, and others. In early 1992, after Smith had left, GM was close to bankruptcy. The lesson is that technology is not a panacea.

Box 1.10 **A note on digitization**

The latest hype in manufacturing is "digitization," "digitalization," or "digital transformation." Digitization, strictly speaking, is the process of converting analog data into digital form. Digitalization is broadly used for the adoption of information technology in organizations. If the transformation currently underway is not just "digitization," then what is it?

Materials are concrete objects: heaps of ores, ingots, slabs, machined parts, motors, etc. Data, on the other hand, is just the first level in a world of abstractions built on top of one another, as follows (the DIKW pyramid):

- Level 1: *Data* is whatever is read or written.
- Level 2: *Information* is what a human or a machine learns from reading data.
- Level 3: *Knowledge* is a collection of information that matches reality.
- Level 4: *Wisdom* is the resourcefulness due to the cumulative assimilation of knowledge.

Computer scientist Don Knuth defined data as "the stuff that's input or output." Our change in vocabulary to "read or written" is to distinguish data from materials, which are also input and output in manufacturing. The above take on information is consistent with mathematician Claude Shannon's, to whom it is, more formally, the reduction in randomness due to reading data. For example, before a final test in an assembly line, you thought a product unit had a 99% probability of being fit for use and passing the final test increases this probability to 99.9999%. If the information is right, you have knowledge about the product state. At the highest abstraction level, wisdom is the ability to use accumulated knowledge to make decisions.

It is clear that digitization alone (that is, making processes create more data) will not do much for manufacturing unless the higher DIKW levels are addressed. The challenge is to transform data to information, knowledge, and ultimately wisdom that can be used to make better decisions.

When we look for a generic term to cover all the applications of computers in manufacturing, IT and OT emerge as the most descriptive and accurate terms.

1.4. **Role of manufacturing today**

Manufacturing is part of the second of three major sectors of the economy. The first sector is agriculture. The third sector is services, which is sometimes further split into classic services such as restaurants, transportation, and policing, and a quaternary sector covering knowledge-based services such as IT services, R&D, consulting, or education. The second sector is misleadingly called "industry." According to the World Bank, industry comprises value added in manufacturing, mining, construction, electricity, water, and gas. The World Bank's definition is congruent with the one we use in this textbook.

The available data on manufacturing employment and output in the world is incomplete, but we can draw the following conclusions:

- Manufacturing matters for the prosperity of nations.
- The share of manufacturing in the Gross Domestic Product (GDP) of advanced economies is holding steady, and the value extracted from it is increasing in absolute numbers.
- Manufacturing's share of the labor force in advanced economies is slowly declining, but not in absolute numbers.

1.4.1. Manufacturing matters

Figure 1.22 plots the contribution of manufacturing versus other sectors of the economy for the 171 countries or economic areas the World Bank provides data for in 2019.[17] To keep small and large economies comparable, the figures are per capita. Manufacturing has a large multiplier effect; an approximate number is that for every $1 spent in manufacturing, another $2 is typically added to the economy in the form of services.

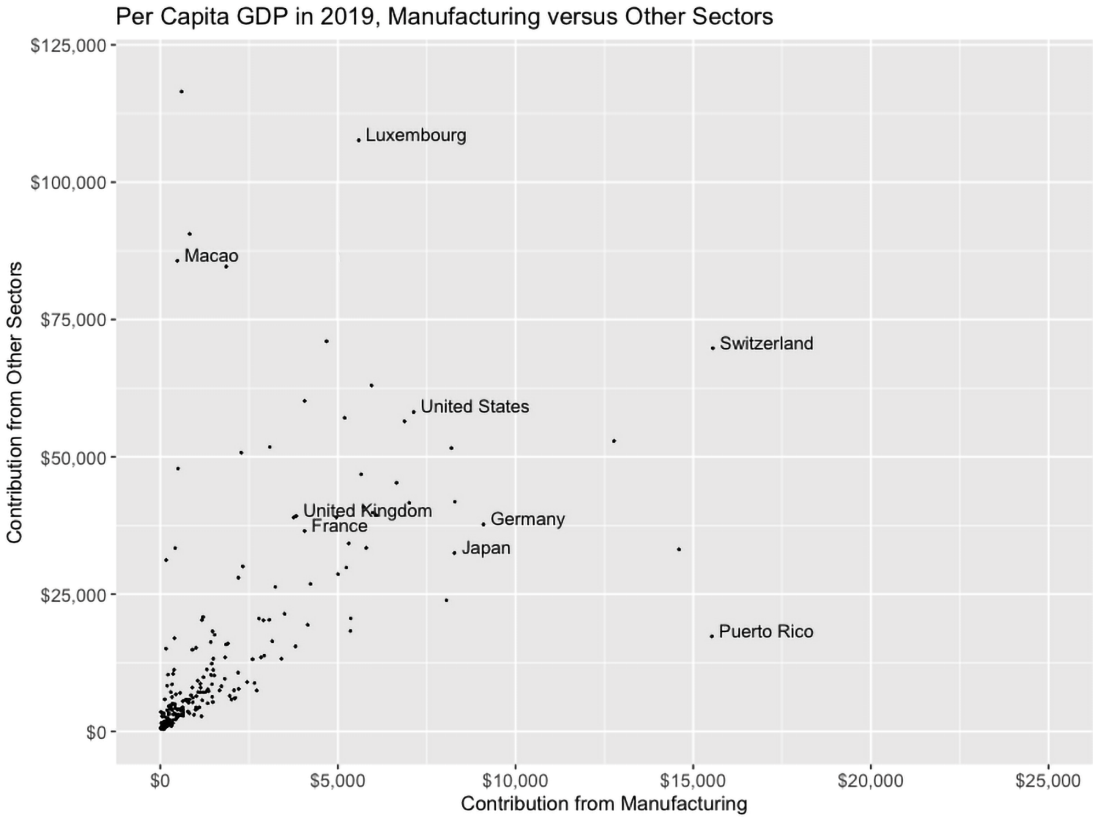


Figure 1.22 In most economies, manufacturing is positively correlated with other sectors

As we can see from Figure 1.22, the size of the manufacturing sector in an economy correlates positively to the sizes of the other sectors. The correlation coefficient is 0.97, indicating that strength in

17 World Bank (2018). World Bank national accounts data, and OECD National Accounts data files. URL: https://data.worldbank.org/

manufacturing and other sectors tend to go together. The outliers with a high per capita GDP and an undersized manufacturing sector both have populations of about 600,000. They are Macao, with revenues centered on gambling, and Luxembourg, a tax haven. The outliers with an oversized manufacturing sector and a low GDP are Ireland and Puerto Rico, respectively, with populations of 5 million and 3.3 million.

> **Stop and Think 1.6**
> Why do manufacturing activities usually generate 2–3 times as many jobs as the direct jobs involved in manufacturing? What are examples of indirectly created jobs?

1.4.2. Contribution of manufacturing to GDP is holding steady

In 2020, manufacturing accounted for about 16% of global GDP and employs nearly one quarter of the global workforce. According to World Bank data, manufacturing typically accounts for around 10–25% of a country's GDP. For most countries, this ratio has been reasonably stable in recent decades (see Figure 1.23). For example, except for a dip during the financial crisis, the share of manufacturing in the GDPs of Japan and Germany has been steady around 22% since 2002. For Germany, 2020 shows a dip due to COVID-19. This share is lower for the UK, France, and the US, at between 9% and 13%, and shows no decline since 2008. It has started to decline in China but is still substantially higher than for the other countries in the group, reflecting the fact that China is still an emerging economy.

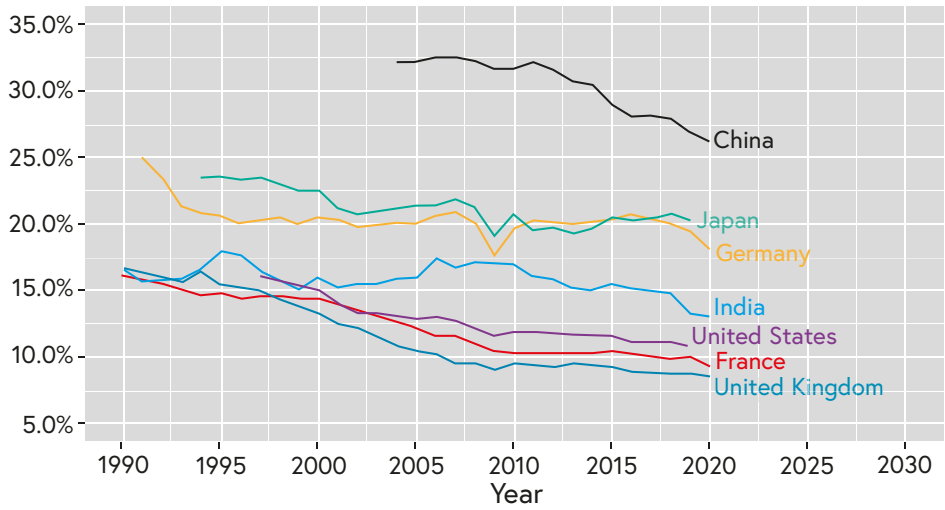


Figure 1.23 Manufacturing's share of GDP is holding steady in the largest economies

Due to the general growth in all the economies in Figure 1.23, the total value derived from manufacturing activities has increased in absolute terms from 1995 to 2020, except for the UK. By this measure at least, most countries have seen no hollowing of their manufacturing sectors. If you go further back in time to the 1990s, you see a decline in manufacturing's share of GDP in Japan, Germany, the US,

France, and the UK, but not in the past 20 years for Japan and Germany, and not in the past 10 years for the US, France, and the UK.

1.4.3. The relative share of manufacturing jobs is declining

Manufacturing's *share of* the labor force, however, has been declining steadily in almost all advanced countries. The decline in manufacturing workforce has at least two different causes:

1. **Productivity increases reduce the need for labor.** Whenever technology becomes available that is more productive than people and is economically justified, managers will buy and use it. For example, steel works today require 15 times fewer employees today than 50 years ago for the same output.
2. **Increasing labor costs leads to offshoring or outsourcing of labor-intensive manufacturing activities.** Due to its higher labor costs, manufacturing in advanced economies has migrated towards products in which labor is a small part of the cost. In making tractors, labor is a small fraction of the cost, making expensive but highly skilled workers competitive in the world market. It is not the case, on the other hand, in sewing mass-market garments and this sector has almost entirely vanished from advanced economies.

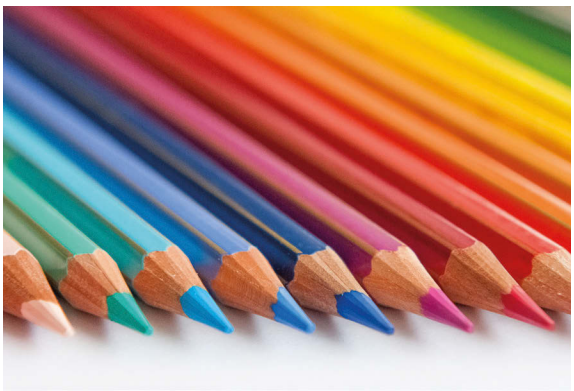


Figure 1.24 The production of pencils by German company Faber Castell is offshored

For example, Faber-Castell, a German manufacturer of pens and pencils, has invested in automation but still only kept R&D and the manufacturing of high-end writing instruments in its headquarters in Stein, Germany, and offshored the production of other pens and pencils to its 14 factories in other parts of the world.

In absolute numbers of people employed, the shift is less dramatic than in the share of the labor force, because of growth in both population and demand, Consider for example the United States. Overall, the US labor force more than doubled between 1960 and 2016. According to the U.S. Bureau of Labor Statistics (2018),[18] *manufacturing* employment was 15.4 million in 1960 and 12.4 million in 2016. The most dramatic reduction happened in the first decade of the 21st century, when manufacturing employment fell from 17.2 million employees in 2000 to 11.5 million employees in 2010. Since 2010, manufacturing employment has been steadily growing in absolute numbers in the US. Its relative share of direct jobs however is just 8.6% in 2016.

As relative employment continues to decline in manufacturing, a challenge in advanced economies is to manage the gradual transition away from manufacturing jobs for most of the population, and the evolution of manufacturing jobs towards higher levels of skills and education. For example, instead of controlling one machine on the manufacturing floor, an employee writes the programs that run three machines automatically. At the company level, managing this transition well is a matter of human-resource policies for competence development.

18 U.S. Bureau of Labor Statistics (2018). *All employees, Manufacturing (MANEMP)* [Data set]. Retrieved from FRED, Federal Reserve Bank of St. Louis. URL: https://fred.stlouisfed.org/series/MANEMP

Box 1.11 **China as "the factory of the world"**

One country stands out in Figure 1.23 on page 38. Over the last half century, China has risen to become the world's second largest economy behind the USA. Ranked by the manufacturing sector's contribution to the Gross Domestic Product, China is number one since 2011. According to data published by the United Nations Statistics Division, China accounted for 28% of global manufacturing output in 2018. One-third of China's total industrial output comes from the coastal provinces of Jiangsu, Shanghai, and Zhejiang around the Yangtze River Delta, and Guangdong around the Pearl River Delta. The latest data cited by the U.S. Bureau of Labor Statistics for manufacturing in China is of 99 million employees in 2009, exceeding all the five other top GDP countries combined. It is not just the massive workforce that makes China a top destination for manufacturing; 168,400 new industrial robots were installed in China in 2020, or 43% of the world's total installations according to the International Federation of Robotics.

China has many advantages for manufacturing. First, its home market size is by far the largest one in the world. Second, China has a political system that provides incentives for investments in manufacturing jobs. Third, an increasingly sophisticated Chinese research and development (R&D) sector has made Chinese companies savvy in IT and new technologies. However, as for any country that is making the transition from a developing to a developed country, China also experiences rising labor costs and skills shortages. As Chinese companies accumulate some of their own intellectual property, they develop a sensitivity to this topic and a respect for the intellectual property of others. Initially, they had nothing to lose; now, they do. Eventually, the Chinese manufacturers started to source from offshore production to other locations. Companies have started to move production to Vietnam, Thailand, Cambodia, India, the Philippines and Indonesia, among other places.



Figure 1.25 More than one-fourth of the world's manufacturing output comes from China

Summary

The most essential points covered in this chapter were as follows:

1. Manufacturing is defined as the transformation of physical materials into physical goods using repeatable processes.
2. Manufacturing takes place in factories and consists of processes, which can be mapped with the input-process-output model.
3. Effectiveness should take priority over efficiency, but both are important for competitiveness.
4. The four Ms are needed simultaneously for a process to run.
5. A systems perspective is needed in manufacturing.
6. Manufacturing is a social activity, organized in departments.
7. Operators are the most essential part of a factory and must be treated accordingly.
8. Manufacturing has continuously developed throughout history. Particular developments in the history of manufacturing were:
   - Craft production
   - Industrialized production
   - The American System of Manufacture
   - Mass production
   - The Toyota Production System and lean manufacturing
   - The introduction of IT and OT in manufacturing
9. Manufacturing is an essential business activity for most economies and contributes a significant share to their GDP (10–30%).
10. Although relative employment in manufacturing is declining in many countries, its share to GDP remains stable with essential spillover effects on other industrial sectors.

Test yourself

1. What is manufacturing?
2. Sketch an input-process-output model for assembling a pen.
3. Why should effectiveness take priority over efficiency?
4. What factors are needed for manufacturing to take place?
   (The MANUFACTURES mnemonic can be helpful.)
5. What is a system, and what does it have to do with manufacturing?
6. Draw a typical organizational chart for a manufacturing factory.
7. How does industrialized production differ from craft production?
8. What was the most striking feature of the American System of Manufacture?
9. What is meant by digitization and digitalization of manufacturing?

10 Use a web search to figure out how much employment and GDP manufacturing contributes within your country. What are the most important manufacturing sectors?

Further reading

Alder, K. (2010). *Engineering the revolution*. University of Chicago Press.

Badiru, A. B. (Ed.). (2013). *Handbook of Industrial and Systems Engineering*. CRC Press.

Chandler, A. (1993). *The visible hand*. Harvard University Press.

Freeman, J. B. (2018). *Behemoth: A history of the factory and the making of the modern world*. W. W. Norton & Company.

Holweg, M., Schmenner, R., Lawson, B., De Meyer, A. & Davies, J. (2018). *Process theory: The principles of operations management*. Oxford University Press.

Hounshell, D. (1984). *From the American system to mass production, 1800-1932*. Johns Hopkins University Press.

Marsh, A. (2018). *The factory: A social history of work and technology*. Greenwood.

Nye, D. E. (2013). *America's assembly line*. MIT Press.

Roser, C. (2016). *Faster, better, cheaper in the history of manufacturing: From the Stone Age to lean manufacturing and beyond*. CRC Press.

2 Manufacturing Strategy

Manufacturing is a strategic activity. Some business executives misinterpret manufacturing as the operating arm of a company; a function supposed to execute strategy, not make it. These companies end up with "millstone" factories, slowly eroding their market position and profitability. As Wickham Skinner[1] argued, manufacturing can be a competitive weapon. Manufacturing as an activity raises strategic questions: How do we serve the demand? Where do we put our factories? How much customizing do we offer? What technologies do we bet on? ... This chapter discusses these questions and the means of answering them.

The manufacturing strategy should follow from business strategy, which should be driven by the company's mission statement. The manufacturing strategy defines how the manufacturer plans to compete in the market. For manufacturing, this means defining priorities and then attempting to develop the corresponding capabilities. While this is much easier said than done, there are many concepts that help us formalize and develop a manufacturing strategy. We introduce some of the most important of these concepts in this chapter.

The issues of manufacturing strategy are best understood through real cases. This chapter is therefore rich in examples. You will learn about elements of the manufacturing strategies of Dell, Honest Tea, IKEA, Intel, Nespresso, Rolex, Toyota, and others. The examples are not intended to tell the complete story of each company but only to highlight some of the strategic choices they made to manufacture their products. The intent is to show the variety of constraints and objectives companies face in different industries and the logic behind strategic choices.

Learning objectives

After working through this chapter, you will be able to do the following:

1. Explain what a manufacturing strategy is.
2. Use the concepts of competitive priorities and capabilities.
3. Discuss Skinner's trade-off model and Ferdows and DeMeyer's cumulative capabilities.
4. Use strategy planning tools such as PESTEL, Porter's 5 forces, co-opetition, clockspeed, and SWOT analysis.
5. Make an informed make-or-buy decision.
6. Use the customer order decoupling point to design an order fulfillment strategy.
7. Differentiate engineer-to-order, make-to-order, assemble-to-order, and make-to-stock strategies.
8. Use the concept of the focused factory.

1 Skinner, W. (1969). Manufacturing: The missing link in corporate strategy. *Harvard Business Review*, *47*(3), 79–91.

DOI:10.4324/9781351110310-3

2.1. **Vision, mission, and goals**

Strategy should be aligned from the "top floor" to the shopfloor, as illustrated in Figure 2.1. If the company is part of a larger corporation, there can also be a "corporate strategy" informing "business strategy." Every company should have a business strategy with a vision, mission, and goals:



Figure 2.1 Strategy pyramid

A **vision** is a thumbnail statement of what the organization intends to become, deliver, or achieve in the future, often stated in competitive terms. Autoliv, a first-tier supplier of safety technology to the automobile industry has the following vision: "Saving more lives."

The **mission** is a concrete translation of the vision, and clarifies the organization's direction and purpose. Autoliv's mission is "Providing world class life-saving solutions for mobility and society." "World class" introduces competition; "mobility," where lives are to be saved.

Although the vision is about the future and the mission about today, there is often overlap. Therefore the vision and mission statement are not always discernible. For example, IKEA (Box 2.1) has the following combined vision and mission statement "To create a better everyday life for the many people."

When a *goal* is specified in detail, it becomes an **objective** or a "SMART goal": *Specific*, *Measurable*, *Attainable*, *Relevant*, and *Time-based*. Manufacturing goals can be to grow revenue, increase capacity, launch a new product, or improve customer satisfaction. Goals are periodically reviewed and renewed.

Functional strategies should be aligned with the business strategy, but not necessarily derived directly from it. There are two sides to strategy formulations: what we *want* to do and what we *can* do. Functional capabilities inform business strategy just as much as the reverse. A company should have a functional strategy for every function: Sales and Marketing, Manufacturing, Supply Chain Management, R&D, IT, Maintenance, Quality, Human Resources, and Finance. Each one of these functional strategies needs its own regularly reviewed goals, broken down into tactics and action plans.

> **Definition**:
> The manufacturing strategy of a company is comprised of the long-term goals for the production of its products and a road map for the production function to help execute the business strategy.

Skinner[2] introduced five decision-making areas that manufacturing strategy is concerned with. In the 2020s, information and data management should be added to his original list:

1. Plant and equipment
2. Production planning and control
3. Labor and staffing

2 Skinner, W. (1985). *Manufacturing: The formidable competitive weapon* (pp. 61–62). Wiley.

4 Product design and engineering
5 Organization and management
6 Information and data management

Box 2.1 **IKEA: Building a lasting competitive advantage**

The furniture giant IKEA, originally from Sweden and now headquartered in the Netherlands, offers 25,000 different furniture and household products through more than 400 stores in 50 countries. It has become the world's largest furniture retail chain and employs 123,000 people. How do they do it?

IKEA's success is to a large extent due to its operations strategy. With the customer in mind, they start with setting the price and then design quality furniture that is easy to manufacture, ship, and assemble. Most of these are stored in flat packs with minimum space waste. They are easy to handle, store, and transport.

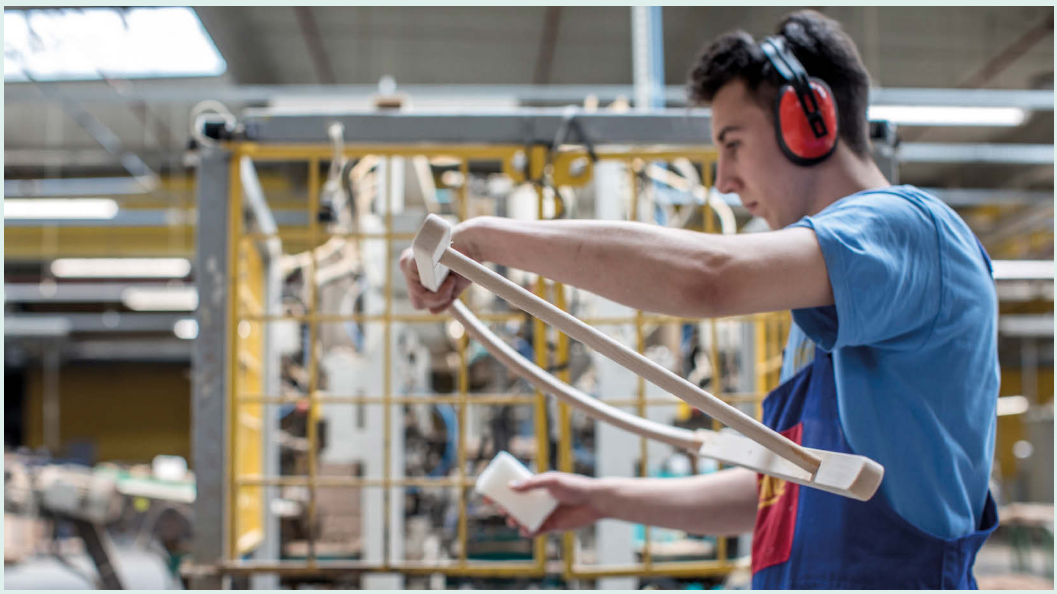


Figure 2.2 IKEA Industry coworker in Romania (Credit: IKEA)

IKEA is strategic about their *make-or-buy decisions*. Around one-tenth of IKEA's retail sales are produced in IKEA's own factories. IKEA set up its own manufacturing branch to reduce the risk of supply disruption and to keep production know-how in-house. This branch, named IKEA Industry, is the largest wooden furniture manufacturer in the world with 20,000 employees in more than 40 factories.

The idea of having customers assemble the product themselves saves costs in manufacturing and transportation. In IKEA warehouses, you get your furniture on the day of purchase, while many other furniture retails chains have long lead times. The days when IKEA furniture were missing parts or assemble instructions were intricate are far gone. In short, IKEA has built excellence in quality, variety, delivery, and cost, and is working diligently on improving sustainability performance.

2.2. **Strategy and tactics**

Both strategy and tactics are originally military terms. Strategy is about picking which battles to fight and tactics about actually fighting them. Originally "the art of the general," from the Greek "strategos" (στρατηγός), meaning "general." It is about which armies or fleets to deploy where and for what purpose. Strategy is a choice of direction, a selection of what an organization wants to achieve. *How* it should go about moving in this direction is tactics. Hence, strategy goes hand in hand with tactics. A strategy without the tactical means to implement it is moot; tactical skills without a strategy have no purpose to serve.

In a multi-level hierarchy, strategy and tactics cascade down from the top to individual contributors. What is tactical to the general manager is strategic to the division managers, what is tactical to the division managers is strategic to the department heads, and so on all the way to production operator. The strategy-making process, however, involves back and forth exchanges between levels. The general needs to know what the troops can do as much as the troops need to know what the general wants them to do. The hierarchical nature of the relationships interferes with the communications. The boss wants subordinates to achieve stretch goals and the subordinates want to please the boss.



Figure 2.3 Strategy is the art of the general making a plan of action

When you make the strategic decision of starting a factory as a greenfield project in a new country, it will be ineffective unless you have a team of people who know the requisite tactics, and it takes specific steps to develop such a team. They must know how to select a location, negotiate with local governments, lay out the plant, set up its supply chain, recruit and train its workforce, etc. You can't just focus on the big picture; you have to dig into the tactical details. Otherwise, you will experience what is meant by the proverb "culture eats strategy for breakfast," a quote often attributed to Peter Drucker.

2.3. **Competitive priorities and capabilities**

Strategy is a managerial decision. Companies *choose* how they want to compete to create value for the customer. It starts with "What does the customer want?" and the follow-up is "What can we possibly deliver?" Based on the answers to these questions, managers develop a manufacturing strategy.

It is common to consider the key attributes of the products and services. Does the customer want high quality, low cost, high safety, fast delivery, innovative products, customized or standard products, a brand identity, ethical and sustainable products? What matters *the most*? The resulting list is matched with what the company *can* and *wants* to deliver and results in its **competitive priorities**.

Ultimately, companies do not compete on the basis of their strategies and tactics – what they want and plan to do – but on their capabilities to carry out the plans – what they do. While some companies are able to turn their **competitive capabilities** into a formidable competitive weapon, others end up with what Skinner called corporate millstones. Some companies compete *despite* poor operational capabilities. Hayes and Wheelwright[3] separated four stages of how manufacturing can contribute to business strategy as shown in Figure 2.4:

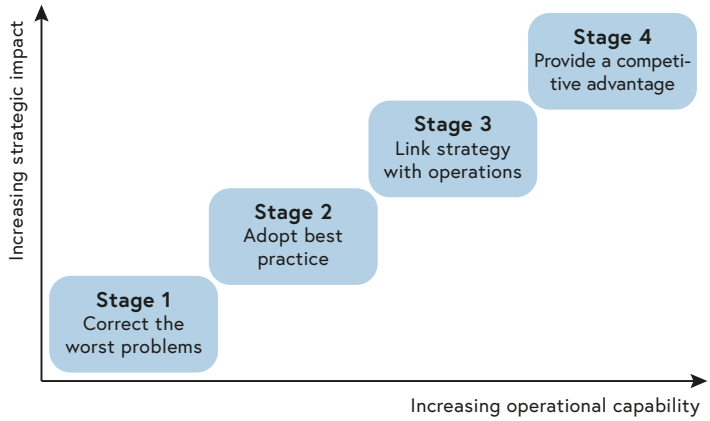


Figure 2.4 The role of manufacturing in business strategy (Credit: Hayes and Wheelwright 1984)

- In stage 1, the company is fighting fires. In some close-to-monopoly markets, companies can do this because customers have no real alternatives.
- In stage 2, companies copy others. This can make them less inept but firms in this stage do not compete on the basis of their manufacturing activities.
- In stage 3, the company links operations with business strategy. In this stage, the firm has become effective and efficient in the areas where it wants to compete.
- In stage 4, the operational capabilities have become a competitive advantage. In these firms, Operations is in the driving seat when strategy is developed. Many of these companies tend to outperform their rivals and become famous, trusted, and valued brands.

3 Hayes, R. H. & Wheelwright, S. C. (1984). *Restoring our competitive edge: Competing through manufacturing*. John Wiley & Sons.

The contribution of operations to performance is not the same in every business. To take an extreme example, Pablo Picasso's studio was a successful business for 7 decades, and its operations consisted of Pablo Picasso doing whatever inspired him at all times, with art dealers dropping by to pick up his output. Success was due to Picasso's creativity, and neither to the organization of the flow of work through his studio, nor to customer service. At the opposite extreme, the manufacturing of products like electrical outlets or scissors is a team effort, against competitors with access to the same equipment under the same conditions, paying the same prices for materials. There, the way operations are designed and run makes the winners and losers.

Most businesses are on a spectrum between these extremes. In luxury goods, for example, operations matter, but they are secondary to brand management. When a diamond-encrusted, inaccurate mechanical watch sells for $5,000, it is not primarily because of the way it is made in the factory. To make a profit selling accurate watches for $10, on the other hand, you have to excel at production operations.

2.3.1. Skinner's trade-off model

The father of manufacturing strategy as an academic subject, Wickham Skinner, introduced a trade-off model that is influential but not universally applicable. It is based on the mistaken perception that you can't combine the best product quality, the lowest cost, and the fastest delivery. This model suggests manufacturers must trade-off between competitive priorities:

- **Low cost versus high quality**. For example, expensive raw materials can be substituted with cheaper ones of lower quality.
- **Fast delivery versus customized products**. For example, a furniture shop offers standard products on display for immediate purchase, but a carpentry offers tailored furniture at longer lead times.
- **Innovative products versus low cost**. For example, brand manufacturers in the pharmaceutical industry can spend around 20–30% of turnover on R&D for new products while generic drug manufactures can mass-produce medicine at low cost without R&D expenses.

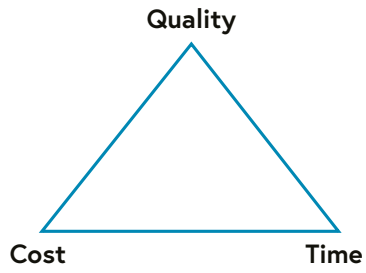


Figure 2.5 The trade-off model

Figure 2.5 illustrates one way to look at trade-offs. It suggests that manufactures cannot excel at cost, quality, and time at the same time. According to proponents of the trade-off model, companies need to pick any 2 of these 3 to compete on. It suggest that any manufactured product can be delivered with low quality cheaply and fast, as in the bottom of the triangle, with high quality fast but expensive, as on the right side, or with high quality, at low cost, but slow, as on the left side.

As we will discuss next, this is *not* always true. Some companies have been successful pursuing all three.

Stop and Think 2.1

Imagine you are planning to start a watch factory, what are driving the possible trade-offs in competitive priorities? What if you were planning to start a pizza restaurant?

2.3.2. **Ferdows and De Meyer's cumulative capabilities model**

Despite trade-offs, some companies seem to excel in many dimensions of competition. Consider for example Dell and IKEA discussed earlier. Another company that is famous for its manufacturing strategy is Toyota Motor Corporation. In the 1970s and 1980s, an M.I.T. led team of researchers participated in the International Motor Vehicle Program – a benchmarking study of automobile production. They found that Toyota's production system outperformed mass production in nearly all performance categories:

> *"Lean production uses less of everything, compared with mass production. Half the human effort in the factory, half the manufacturing space, half the investment in tools, half the engineering hours to develop a new product in half the time. Also, it requires keeping far less than half the needed inventory on site, results in many fewer defects, and produces a greater and ever-growing variety of products."*[4]

As is clear from this quote, Toyota competed not only on low cost, but also on high quality, short delivery lead times, and high product variety offering. How was that possible?

The *cumulative capabilities model* by Kasra Ferdows and Arnoud De Meyer[5] suggests that they can do it by paying particular attention to the sequence in which capabilities are built. This model is also known as the "sandcone model" due to its shape and logic. If sand represents resources, the model posits that building cumulative capabilities is like building a sandcone: You need exponentially more sand for every additional inch. But crucially, most of the resources would need to go to building a wider base. You build the capabilities on top of each other – not by trading off one against another. Note that improving *quality* is at the base.

As illustrated in Figure 2.6, building higher-layers capabilities requires continuous investments in the lower layers. This way, manufacturers build capabilities *cumulatively*. The net result of improving capabilities cumulatively is a better and more lasting cost efficiency.

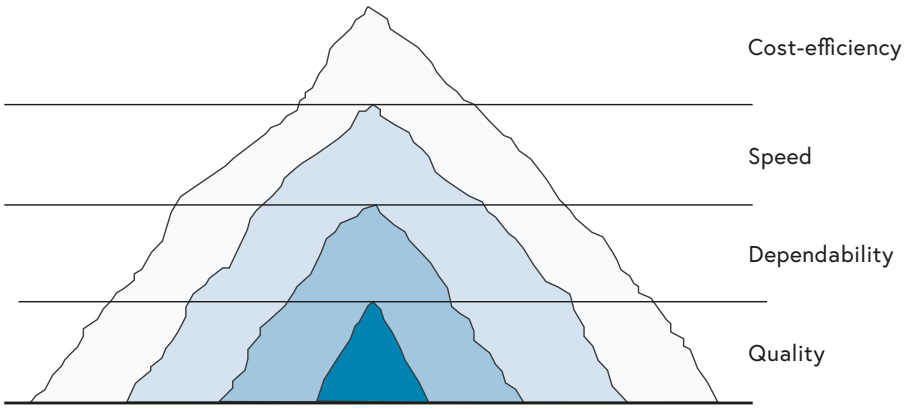


Figure 2.6 The sandcone model for cumulative capabilities (Credit: Kasra Ferdows and Arnoud De Meyer)

4 Womack, J. P., Jones, D. T. & Roos, D. (1990). *The machine that changed the world*. Rawson Associates.

5 Ferdows, K. & De Meyer, A. (1990). Lasting improvements in manufacturing performance: In search of a new theory. *Journal of Operations Management, 9*(2), 168-84.

2.4. Strategy planning frameworks

Some strategic planning tools are helpful for manufacturers to keep in mind when developing a manufacturing strategy. Three of the most used ones are PESTEL analysis, Porter's five forces, and SWOT analysis. These frameworks are so intuitive that a short introduction suffices here. There are also plenty of detailed explanations and examples on the internet.

2.4.1. PESTEL analysis

PESTEL is an acronym for Political Economic, Social, Technological Environmental, and Legal, which represent categories used to analyze a markets or countries. A PESTEL analysis covers the macro-environmental factors that may impact an organization's performance. The key to a good PESTEL analysis is not to map a long list of factors but to focus on key issues for the company. The abbreviation PESTEL has developed in several iterations and exists in many forms but its origin is often traced to Francis Aguilar's ETPS scanning tool.[6]

Used for monitoring, a PESTEL analysis can inform the manufacturer about macro-trends affecting products and processes. New environmental regulations may forbid the use of materials or restrict factory emissions; new technology may enable new processes; economic changes may affect the price of materials. A PESTEL analysis should be carried out when planning new factories or warehouses in foreign countries.

Table 2.1 lists some of the most common factors considered in a PESTEL analysis but there are many more.

Table 2.1 PESTEL analysis with some usual factors to consider

P	E	S	T	E	L
Political	**Economic**	**Social**	**Technological**	**Environmental**	**Legal**
Government stability Tax policy Regulation Infrastructure	Economic growth Inflation Monetary policy Employment rates	Demography Education Lifestyle factors Income distribution	Technology development R&D initiatives Connectivity Technology transfer	Regulations Climate change Energy policy Workforce health	Regional laws Court system Health and safety Law enforcement

6 Aguilar, F. J. (1967). *Scanning the business environment*. Macmillan.

Box 2.2 **Dell: Competing through velocity**

Dell is now the 3rd largest manufacturer of PCs, with a 17% share of market, behind Lenovo and HP. Dell receives custom orders for computers and peripherals online, buys the components, assembles, tests, and ships the units. The selling price of the products is barely 10% higher than what Dell pays for components. However, Dell is able to collect money from the customers before they have to pay their suppliers. This positive cash flow cycle has propelled Dell to its position on the market share podium.



Figure 2.7 Dell building
(Credit: Dell Technologies)

Dell buys integrated circuits from semiconductor companies. On the other hand, the level of assembly work done by Dell, at least for desktop computers, is well within the capabilities of do-it-yourselfers. A middle school student following online instructions can assemble a desktop computer from a kit including a case, a motherboard, a CPU, a drive, etc., in about one hour. So why do customers prefer to buy from Dell or one of its competitors?

It's for the support. After assembling each computer, Dell tests it for 12 hours and takes some responsibility for fixing it if it breaks. If, on the other hand, you assemble your own computer, you are on your own. If, every week, you buy $1,000 of components, assemble them into a computer and sell it for $1,100, after 52 weeks, it will add up to $5,200, or more than 5 times the amount spent on components. If, on the other hand, it takes you 3 months to sell the computer, you repeat that 4 times in a year and get $400. In other words, in this business, the only way to make money is to have short lead times.

In addition, in the computer business, the price of a given component drops by about 6% per month. In other words, if you have bought the components you use today a month ago, you paid 6% more than if you had bought them today. A company like Dell, therefore, seeks to both buy the components as late as possible and deliver the goods as fast as possible. They refer to their performance in this area as velocity.

The short product life is another characteristic of this activity; 90% of the sales volume is for products introduced in the past 9 months. This makes investing in product-specific production equipment unattractive. The computer market is also subject to structural changes. According to Statista, in 2010, the world market for computers was 157M desktops, 201M laptops, and 19M tablets; in 2020, 73M desktops, 198M laptops, and 150M tablets. In short, holding physical infrastructure is a risk. Today, almost all PC assembly is done by contract manufacturers that assemble computers for a variety of labels. By 2020, Dell's laptops are assembled by Taiwanese companies Quanta and Compal, while Dell still assemble their own desktop PCs and servers in seven factories worldwide.

Box 2.3 **Toyota's powerful manufacturing strategy: The Toyota Production System**

The Toyota Production System (TPS) represents Toyota's manufacturing strategy. It is common to represent TPS as a house as shown in Figure 2.8, which must be built ground-up. The foundation of TPS is operational stability, which is needed for effective and efficient production. On top of this foundation are the two principles of just-in-time and jidoka.



Figure 2.8 A representation of the TPS House

Just-in-time means always delivering the right parts at the right time in the right quantities, with shortage prevention based on vigilance rather than inventory.

Toyota's *jidoka* is their unique, incremental approach to automation focused on human–machine systems, which has, since early days, included automatic stopping mechanisms when equipment malfunctions. It is often mistakenly reduced to "built-in-quality."

These two key pillars of the TPS were developed by members of the Toyoda family. Before Toyota made cars, they made looms. Toyota Loom founder Sakichi Toyoda developed looms that automatically switched shuttles and stopped when the weft broke. This approach later grew into *jidoka*. Later, Kiichiro Toyoda developed just-in-time. In the aftermath of World War II, Toyota could not afford truckloads of parts and found ways to produce by replenishing the quantities consumed on the cars it sold.

While having these key principles is useful, factories also need to implement and sustain them. Therefore, Toyota works continually to improve performance, which is known as *kaizen*. For example, over many years, Taiichi Ohno experimented with quick changeover of dies of car body presses. Toyota eventually reduced the die exchange time from 1 day to 3 minutes. Instead of changeover specialists, the die exchange could now be done be by operators. This made it cheaper to run small batches than the large batches others used.

It enabled just-in-time production and eliminated the need to carry large inventories. But, more importantly, it made stamping mistakes almost immediately visible. It enabled Toyota to produce higher quality cars at a lower cost and faster than its competitors. The TPS enabled Toyota to defy the trade-off model that its competitors took as granted.

What is not apparent in the TPS house, in Figure 2.8, is the essential role of people development in Toyota. Toyota's slogan, "Good Thinking, Good Products," won an in-house slogan contest in 1953 and can still be found in all Toyota factories as a poster or banner. It suggests that for a company to make good products, it first has to develop its people. The importance of people in Toyota are covered in later chapters.

2.4.2. **Porter's five forces**

Michael Porter's five forces model is shown in Figure 2.9. It consists of five "forces" that affect the competitive rivalry of a business.[7] First, located in the center is the existing rivalry among competitors. This rivalry is dynamic, which Porter illustrated with the revolving circle. Four other forces can alter this rivalry: The bargaining power of suppliers, the bargaining power of buyers, the threat of new entrants, and the threat of substitute products or services.

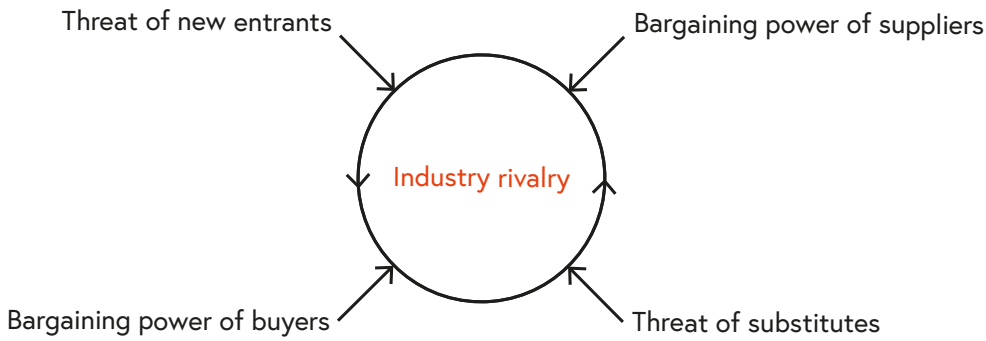


Figure 2.9 Porter's five forces model

Note that Porter's five forces model takes an outside-in perspective. It is essentially an analysis of market forces, which of course is also important for manufacturers. For a manufacturer to gain a competitive advantage it needs to understand the forces that affects its competitive position. All the five forces are external threats that should be monitored. Porter's model provides a simple and popular template, which can work as a starting point for deeper analysis.

2.4.3. **Brandenburger and Nalebuff's co-opetition**

In 1996, Adam Brandenburger and Barry Nalebuff introduced the concept of co-opetition as a game a company plays with suppliers, customers, competitors, and *complementors*, a new term they coined.[8] Figure 2.10 shows the main categories of players a company interacts with.

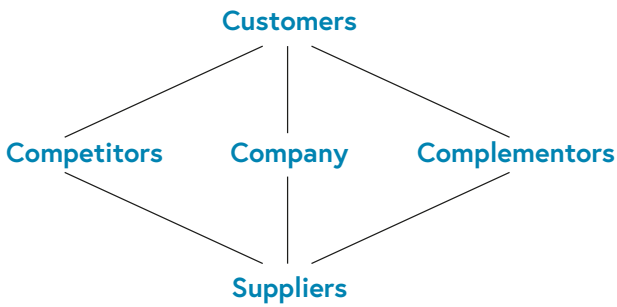


Figure 2.10 The co-opetition model

7 Porter, M. E. (1979). How competitive forces shape strategy. *Harvard Business Review 57*(2), 137–145.

8 Brandenburger, A. & Nalebuff, B. (1996). *Co-opetition.* Currency/Doubleday.

In this game, the value added by a player is the amount each player contributes to the pot that is eventually shared:

- **Competitors** are defined as players whose presence *decreases* the company's value added by providing customers with alternatives to using the company's products. In this broad sense, railroads and cargo planes are competitors to trucks.
- **Complementors** are players who *increase* the company's value added by making its products more useful to customers. Examples include networks of gas stations for car makers and software developers for computer makers.

The title of the book, *Co-opetition*, comes from behaviors observed in many types of businesses, where competitors also collaborate, for example by setting up shops at the same locations, like the diamond district in Antwerp, Belgium, the Akihabara for consumer electronics in Tokyo, Napa Valley for California wineries, or Silicon Valley for computer technologies.

In all these cases, competitors collaborate to nurture the reputation of the area, and then compete against one another within it. As recounted by George Taber, the Napa Valley wineries caught up with their French rivals by sharing the results of their experiments in wine making. When Silicon Valley chipmakers develop a new process, they tell their competitors about it, in the hope of making it a standard that equipment manufacturers will build machines for, at a cost shared by multiple users. They then compete on how well they are able to execute this process.

At the 2019 TOCICO conference in Chicago, Boeing executive Steve Holt introduced a speaker from competitor Embraer with the following words: "Companies in our business have been competitive but we have also worked together to make aviation as a whole a better place." This is typical co-opetition.

2.4.4. **Fine's clockspeed**

Some products have lives in decades and others in months. The Boeing 737 has gone through many versions since it entered service in 1968 but some of its components remained unchanged for decades. In businesses like laptop computers or smartphones, product lives are under one year due to technology. In toys, it may be as short as a Christmas season.

In *Clockspeed*,[9] Charles Fine analyzes the effect of product life on manufacturing strategy. With short product lives, managers experience the ultimate consequences of their decisions while still in the position they had when they made them. With product lives in decades, they may be retired or dead long before these consequences unfold. Outsourcing components or licensing production today may nurture a competitor 30 years from now. Promising generous retirement benefits to a young workforce may bankrupt the company when this cohort eventually retires.

Setting strategy for a company in a low clockspeed industry requires managers to care for what will happen to it long after they are gone. In particular, it may *require* them to make decisions that sacrifice short-term profits.

Box 2.4 illustrates the challenges faced by the high clockspeed toy industry during Christmas season.

9 Fine, C. (1998). *Clockspeed: Winning industry control in the age of temporary advantage.* Perseus.

Box 2.4 ***Jingle all the way* is difficult in high clockspeed industries**



Figure 2.11 It is difficult to fulfill all demands

Jingle all the way is a 1996 movie about a father trying to buy a popular toy for his son on Christmas Eve. Most toys in the West are sold before Christmas each year, and the bulk of these toys are made in the Pearl River Delta of China from June to August, before the manufacturers can know which ones will be popular. This information is only known after the entire Christmas supply has been made, packaged, and shipped. As a result, every year there is a shortage of popular items and a glut of duds.

Ideally, a local factory should be able to ramp up and make more of the popular items during the Christmas season. The popular items, however, may be anything from an action figure like the movie's "Turbo Man" to a Cabbage Patch Doll or a game console. To this date, no manufacturer has been able to build a factory with the agility required for this kind of range.

2.4.5. **SWOT analysis**

A SWOT analysis maps the internal **strengths** and **weaknesses** of a company and its external **opportunities** and **threats**. The purpose of a SWOT analysis is to conduct a general and quick examination of the current status to help the company develop a strategy. The result of the analysis may provide a basis for a more detailed analysis in certain areas. The SWOT analysis framework in Table 2.2 is self-explanatory. It can be used with post-it notes on a board, pen and paper, or slideware during a workshop.

Table 2.2 SWOT analysis framework

S TRENGTHS	**W** EAKNESSES
O PPORTUNITIES	**T** HREATS

Stop and Think 2.2
Make a quick SWOT analysis of yourself as a learner. Use it to develop a learning strategy for *Introduction to Manufacturing*.

Box 2.5 **Intel: Staying at the cutting-edge of technology**

Intel's complex in Rio Rancho, New Mexico, USA, is among the largest semiconductor manufacturing sites in the world. It covers 180 acres for 3,000,000 square feet of manufacturing and office space and employs 1,800. The site grew over decades, along with a town in a previously deserted area North of Albuquerque, as a result of a manufacturing strategy driven by specifics of the semiconductor industry. It is characterized by sequences of hundreds of fabrication operations turning a wafer of silicon with a unit cost on the order of a restaurant meal into a set of integrated circuits worth the price of a luxury car.



Figure 2.12 Photolithography processes in an Intel fab in Oregon, USA (Credit: Intel Corporation)

In the semiconductor industry, a factory is called a "fab" (from semiconductor fabrication). In 1980, Intel bought a 180-acre (73 ha) plot of desert and populated it with several generations of wafer fabs over 40 years, moving from North to South. Raw land in the desert is relatively cheap. Intel bought a plot comparable in size to a car assembly plant making 300,000 cars per year, but initially built on it a facility (Fab 7) that barely occupied one tenth of it. Designed in the late 1980s to make microprocessors, Fab 9 was built as a shell to be populated over time with wafer fabs of several generations, called Fab 9.1 through 9.4. Today, the site is full.

At $2 billion per plant, semiconductor wafer fabrication factories are among the most expensive factories in the world. Within the buildings, the cleanrooms where fabrication takes place accounts for a small fraction of the building space. Fab 11E has 200,000 square feet of class 10 cleanroom space, for 733,000 square feet of manufacturing space, corresponding to the cleanroom occupying the center floor of a 3-story building. In addition, the manufacturing building is connected by a walkway to an office building of 110,255 square feet of footprint and multiple floors. In a semiconductor wafer fabrication plant, the land, buildings, and facility systems account for 12% of the investment; 88% goes to process equipment, which is obsolete in four years. Afterwards, it can be used for many years through the tail of product lives, while state-of-the-art processes require new machines. For these reasons, Intel bought land and built facilities years before they are needed. Buying cheap land in 1980 was a strategic decision.

2.4.6. A resource-based view on manufacturing strategy

Manufacturers have to compete with the resources they have. For this reason, it is important to always keep in mind the capabilities of a manufacturing business when developing its strategy. What a manufacturer can do is constrained by its resources. A resource-based view on strategy should therefore complement the market-based perspectives. The "strengths" and "weaknesses" in a SWOT analysis touches upon internal resources but a more detailed understanding of resources, capacities, and capabilities are needed for developing a viable manufacturing strategy.

By definition, a business is a collection of technical, financial, and human resources brought together and organized for an economic activity. Within this broad context, there are a couple of peculiarities to a manufacturing business. In particular manufacturing is characterized by the magnitude of the *technical* resources needed in the form of facilities, machines, processes, and materials, and the required combination of engineering, management, and operational skills needed for success.

The economic activity of a manufacturing company is making specific goods to sell in a specific market. When markets change, trading companies can easily pivot to sell other goods in other markets. A manufacturing company, on the other hand, is constrained in the range of changes it can make.

The idea that companies compete with the resources they possess is captured by the resource-based view of the firm.[10] This theory is concerned with how the company's assets, resources, and capabilities can be developed to improve the company's performance and competitive position, and ultimately yield a sustainable competitive advantage. Robert Grant has presented a resource-based model that is well aligned with the needs of manufacturers (Figure 2.13).[11] The resource-based model starts with what an organization can do, and seeks to develop the capabilities needed to compete in the market. This model complements the externally oriented PESTEL and Porter's five forces models.

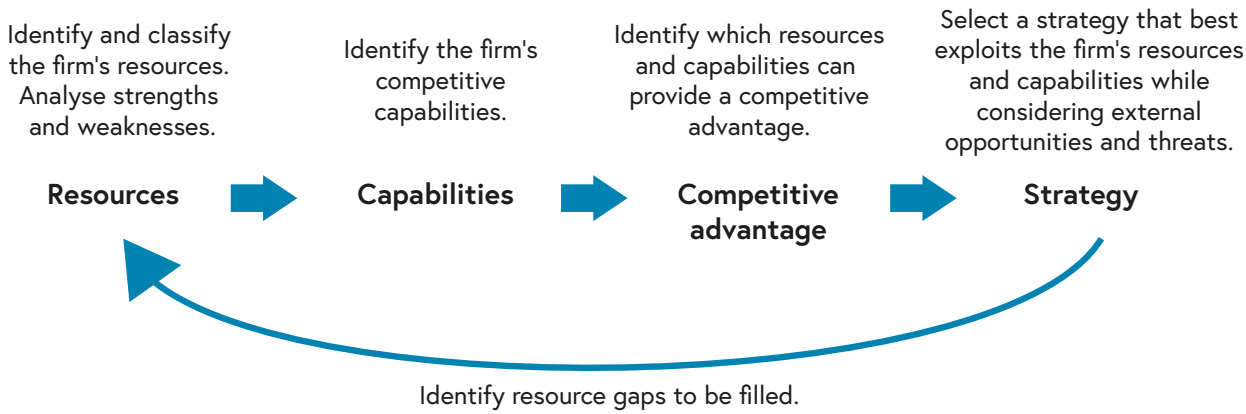


Figure 2.13 A resource-based view on manufacturing strategy

10 Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120.

11 Grant, R. M. (1991). The resource-based theory of competitive advantage: Implications for strategy formulation. *California Management Review*, Spring, 114–35.

2.5. **Make or buy**

A critical part of manufacturing strategy is the determination of whether to produce or buy parts and products. This is referred to as the *make-or-buy decision*. If the manufacturer decides to buy a product or part, it is called *outsourcing*.

A company might outsource any of the following manufacturing related functions:

- Designing the product
- Purchasing the raw materials
- Processing the subcomponents, subassemblies, major assemblies, and finished product
- Storing the products
- Distributing the product

Many companies outsource support functions such as:

- Payroll
- Billing
- Ordering
- Website development and curation
- Employee recruitment
- Facility maintenance
- Food services
- Security
- Janitorial services
- Landscaping

In some manufacturing industries it has become usual to outsource almost everything except the development, design, marketing, and sales of a product. Since the 1990s, the world has seen the rise of contract manufactures such as Taiwan-based Hon Hai Precision Industry (now Foxconn), Singapore-based Flex International Ltd., US-based Jabil Circuit, and Luxembourg-based Elcoteq. Today, these companies are producing, by far, the largest market share of consumer electronics and a growing share of medical equipment, white goods, and a range of other products. Why? Because contract manufacturers offer a manufacturing capacity and agility unlike any original equipment manufacturer. By pooling products from different companies they achieve both economies of scale and economies of scope.

Another reason for massive outsourcing of manufacturing operations is that the perceived value of doing in-house manufacturing is low. Acer founder Stan Shih suggested the "smiley curve" in 1992, which aims to explain where companies can add more or less value to a product (see Figure 2.14). According to Shih, manufacturing is the activity that adds the relative lowest value in the value chain from R&D to after-sales service provision.

Rather than a universally valid model, the smiley curve can be used as a framework for reflection. Some companies have taken the model at face value. Acer, for example, used it to reorient itself from a manufacturing company into a brand house for computer-related products and services.

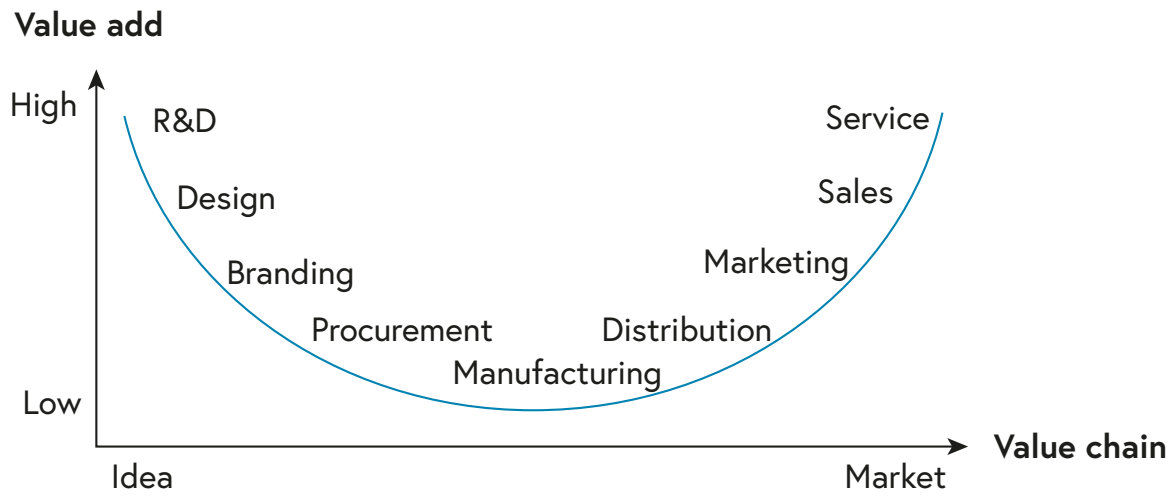


Figure 2.14 Stan Shih's smiley curve for value added

Or consider Apple Inc. Apple designs iPhones in California in the USA and operates several R&D sites across the world. Apple also maintains its own vertically integrated sales and after-sales service network. The manufacturing of iPhones, however, is entirely outsourced to partners, like Foxconn, that assemble the phones in China, Taiwan, and India. Note that Apple maintains tight control of the assembly processes of its contract manufacturers. All parts that go into an iPhone are purchased from suppliers but the operating system is not.

Other companies have chosen different strategies. German teddy bear manufacturer Steiff had outsourced parts of the manufacturing to Chinese suppliers in 2004 but reconsidered this in 2008, after experiencing both quality issues and long lead times. Today, Steiff produces in its own factories in Germany, Tunisia, and Portugal. At one extreme is Rolex, which today is operating a fully integrated and quite secretive supply chain (see Box 2.6)

Box 2.6 **Rolex: A vertically integrated manufacturer**

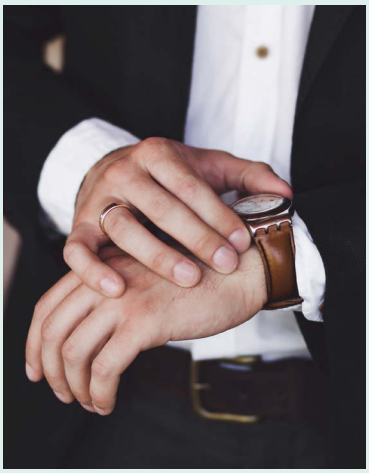


Figure 2.15 A luxury watch

Rolex is perhaps the most famous watchmaker in the world. The company, based in Geneva Switzerland, invented among other things the first waterproof wristwatch in 1926 and the perpetual self-winding rotor in 1931. Production of luxury mechanical watches requires an attention to detail and precision unlike most other products.

In 2004, Rolex bought a watch movement factory for ~$1B. This completed the vertical integration of Rolex, which now produces every part of the watch in a supply chain consisting of four factories located in Switzerland, including Rolex's own gold and steel foundry. Rolex does not only manufacture its watches but also the machines and tools they use to make them.

Box 2.7 **Honest Tea: Make or buy?**

In 1997, Seth Goldman partnered with Barry Nalebuff, his former Yale professor, to start a beverage company they called Honest Tea. By 2017, it was a division of Coca-Cola with $222M in annual sales. In 2013, in collaboration with illustrator Sungyoon Choi, they published Mission in a Bottle, a comic book account of the history of Honest Tea, describing, in particular, the various approaches they took to brew and bottle their beverages in increasing quantities, in other words, their manufacturing strategy.



Figure 2.16 A typical bottling factory: Should Honest Tea make or buy their juice?

When demand outgrew the capacity of Goldman's own kitchen, they outsourced production to a contract manufacturer, which eventually they could not count on. During apple harvest season, for example, their co-packer dedicated its capacity to apple juice and cider and stopped work for Honest Tea. They therefore bought a bottling plant, to control their production capacity and protect their proprietary information. This, however, drove up their fixed costs and increased their vulnerability to changes in markets. Then they involved Coca-Cola, first as a 40% investor in 2008 and then as a parent company in 2011, and Honest Tea products are now made in Coca-Cola factories.

To build Honest Tea as a business, Barry Nalebuff and Seth Goldman faced many challenges besides manufacturing, including product design, marketing, distribution, human resources, and financing. As entrepreneurs, they had to address all of them. As of 2020, Barry Nalebuff still teaches at Yale School of Management and has founded two other companies in the food industry, Real Made and Kombrewcha. Seth Goldman is chairman of Beyond Meat, a maker of vegetable-based protein products, and founder of PLNT Burger, a vegetarian restaurant chain.

2.5.1. Motivation to make or buy

What makes some companies like Rolex and Steiff decide to manufacture in-house, while others like Apple and Acer decide to outsource this essential part of any product supply chain? The make-or-buy decision has been subject of active research for over a century. The behavior of companies reveals that it is often not based on short-term economics. Companies *buy* to reduce costs, particularly the fixed costs of factory ownership; they *make* to control the supply.

Honest Tea (Box 2.7) first made beverages in a kitchen, then outsourced when demand outsized production capacity, then built its own plant to make sure the capacity was available and to prevent their secrets leaking, and finally sold the company to Coca-Cola. When Boeing acquired the Iron Fireman company of Portland, OR, in 1974, it drove up the costs of the parts but it ensured that the machining capacity and know-how were exclusively available to Boeing. Toyota's policy on many components is to develop an in-house capability while outsourcing most of the production.

Buying, however, is an option only when suppliers exist. In a nascent industry, they don't. Semiconductor manufacturers in the 1960s had no choice but to pull their own silicon crystals from sand and saw these crystals into wafers. They even had to build their own production machinery. By the 1980s, they had discontinued in-house crystal pulling and equipment operations. Instead, they could rely on a supplier base that had pulled ahead of them in technology and performance.

In mature industries, geographical locations may make buying impractical even when suppliers exist. There is, for example, an incentive to refine low-grade ores into the desired materials close to the mines instead of shipping raw ore that is 98% waster. Mines, however, are often in remote locations where external services are not available.

2.5.2. The core competency theory

In 1990, C. K. Prahalad and Gary Hamel[12] introduced the concept of core competencies as a basis for make-or-buy decisions. The strategy is to outsource all activities except those based on its "core competencies" or "core technology." What these core competencies are, however, is not always easy to identify. Some cases are straightforward, such as the following:

- If you are assembling airplanes, then feeding lunch to employees is not a core competency, doing it internally is a distraction, and most companies use catering services instead. They only operate their own restaurants where such services are unavailable.
- Intel is turning raw silicon into microprocessors, using multibillion-dollar wafer fabrication plants, and may well consider that inbound and outbound logistics are not part of its core competency while wafer fabrication is.
- Dell assembles custom computers in high volume with short order-fulfillment lead times. Logistics is obviously a core competency for Dell, as is widely recognized.

Others are not so obvious. For example, most carmakers count stamping floors, hoods, doors, and other large components of car bodies among their core technologies. But what about the use of small presses on the side to make bracketry? It can be argued that it is an additional application of the same

12 Hamel, G. & Prahalad, C. K. (1990). Core competence of the organization. *Harvard Business Review, 68*(3), 79–91.

technology and should stay in-house, or that it distracts from the focus on large body parts and should be outsourced.

A "core technology" should be a skill at which no competitor outperforms you. Practically, this may be impossible to know. In addition, even if a manufacturer is underperforming at every individual process, it can outperform competitors by integrating the operations better. Even if you establish that logistics is a core competency, it does not necessarily follow that you keep it in-house. The logistics expertise of Toyota or Dell does not prevent them from outsourcing many logistics-related activities and a substantial part of their expertise is in fact in the management of external providers.

In the absence of a clear definition of a core technology or a core competency, it ends up being whatever management decides to keep in-house. In other words, more specific, operational criteria have to be used to make these decisions. Box 2.8 shows an example from Nespresso.

Box 2.8 **Nespresso: Make, buy, or ally?**

Consider Nespresso, owned by Nestlé. Nespresso is a coffee machine that brews espresso-like coffee from single-serving aluminium capsules. What is Nespresso's core competency? What parts of its products should Nespresso make, and what parts should it buy? For the parts bought from suppliers, should Nespresso form a strategic alliance or rely on arm-length relationships?



Figure 2.17 Aluminium capsules with espresso coffee (Credit: Nespresso)

Nespresso grinds coffee in three factories in Switzerland using beans from contracted farmers working in an alliance with Nestlé (see Nespresso AAA Sustainable Quality Program). For the capsules, however, Nespresso has allied with two manufacturers of aluminum capsules. Having two competing suppliers enables Nespresso to keep price and quality pressures high. It also provides resilience in case of an incident in one of the manufacturers that impairs deliveries.

Nespresso coffee machines are designed but not made by Nestlé. They are produced by third-party manufacturers in strategic partnerships (for example, Eugster/Frismag producing brand names like DeLonghi, Krups, and Magimix).

The cardboard packaging material and a diverse range of capsule containers and holders are bought from suppliers.

2.5.3. **Transaction cost theory**

Once you have decided that you wish to outsource a particular good or service, you still need to validate the economics of doing it. The most significant theoretical framework that assists manufacturers in making this critical decision is *transaction cost theory*, also known as "transaction cost economics," which helps determine whether to make or buy case-by-case.[13]

The theory assumes that people are rational opportunists. According to economist Oliver E. Williamson,[14] there are two types of costs that decide whether an object should be made or bought: the *transaction costs* of buying it and the *production costs* of making it.

Transaction costs include the costs of searching for vendors, processing the transaction, risk hedging, control, and follow-up. The decision to make or buy is based on a comparison of the costs of buying versus making. Williamson defined three factors that impact transaction costs:

- **Transaction frequency**. The traditional argument is that, the more often a transaction is conducted, the higher its costs, and the more attractive in-house production becomes. Hence, high transaction frequency suggests a make-decision.
- **Asset specificity** is a characteristic of the resources used in a transaction. It is high for a technology or resource with little value outside the relationship. If the resources involved in producing the product or service are not easily used for anything else, then suppliers are likely to increase the transaction costs in order to hedge against risk. Moreover, suppliers' production costs are likely to be higher if they must invest in resources that cannot be employed in other transactions. Therefore, higher asset specificity encourages a make-decision.
- Uncertainty in a transaction has two dimensions:
  - The **environmental uncertainty** is the degree to which deliveries can be planned in contracts.
  - The **behavioral uncertainty** is the degree to which actual deliveries can be measured and controlled.

  The traditional transaction cost argument is that high uncertainty of either type encourages a make-decision.

Production costs are the costs related to making the product in-house. How these should be measured is a never-ending discussion in the accounting literature. The traditional, 100-year-old method used by many manufacturers is to obtain a unit cost by adding materials and labor, and allocate overhead based on labor. While widely recognized as inadequate by today's management accountants, it has yet to be systematically replaced.

A manufacturer is not making a single unit but a flow of units over time, and the cost of deciding to do it in-house is the *schedule* of outflows that it triggers, to be compared with the schedule of outflows generated by buying from a supplier. These schedules then lend themselves to the net present value (NPV) analysis that all MBA students learn, and this analysis must be done on multiple scenarios about future demand.

Transaction cost theory quickly becomes complicated. In most cases, simpler derivative frameworks can assist the decision to make or buy. One example is the make-or-buy decision matrix in Figure 2.18.

13 Coase, R. H. (1937). The nature of the firm. *Economica, 4*(16), 386-405.

14 Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. Macmillan Publishers.

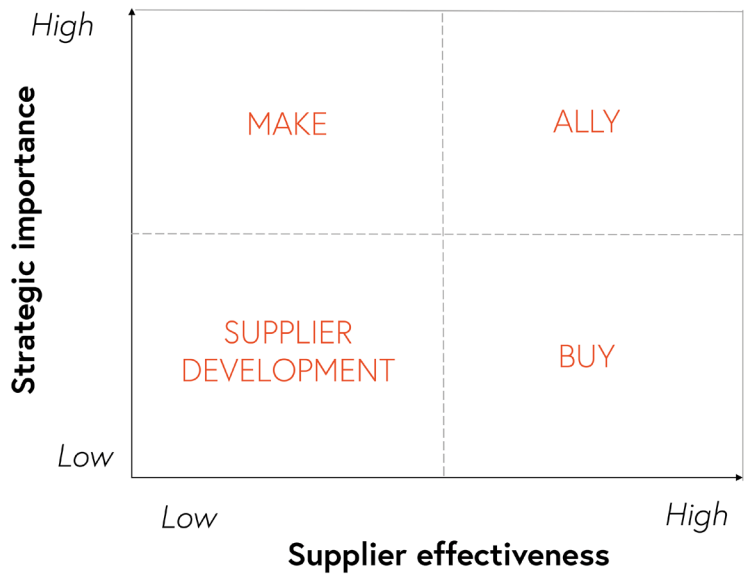


Figure 2.18 A make-or-buy decision matrix

This figure splits strategic importance along the vertical axis and supplier effectiveness along the horizontal axis. The framework suggest that if a product or service is in the upper left quadrant, it is of high strategic importance and suppliers are not effective in supplying it, make it. If it is in the bottom right quadrant, it is of low strategic important and suppliers are effective, buy it. If it is of low strategic importance but suppliers are not effective, develop suppliers into becoming effective – that is, move it from the bottom left to the bottom right quadrant. Finally, if it is of high strategic importance and suppliers are effective, make an alliance with suppliers.

Stop and Think 2.3
Use the framework in Figure 2.18 to evaluate Nespresso's make or buy decisions. Do you think the current set-up is a good strategy?

Whatever method is used to decide whether or not to make in-house or buy from suppliers, it remains one of the most strategic decisions a manufacturer makes. It affects the performance of all competitive capabilities and bottom-line financial results. No wonder, the make-or-buy decision is considered one of the most complex and critical decisions managers are confronted with.

2.6. **Strategies to serve market demand**

If you have a 5-day order fulfillment lead time and it takes less than a day to make the product from start to finish, as in computer assembly, you can schedule production directly from the incoming flow of orders. On the other hand, if you have a 3-month manufacturing lead time, as for semiconductors, you can only fulfill orders with a 5-day lead time from inventory, located at a point in the process where you can finish quickly enough. It's called the *customer order decoupling point* (CODP).

> **Definition**:
> The customer order decoupling point separates order-driven activities from forecast-driven activities

The CODP is a point of strategic importance. Upstream from the decoupling point, activities are driven by forecasts. Downstream from it, they are driven by customer orders. There are four strategies for placing it:

- **Engineer-to-order (ETO)**. Products are engineered and designed to customer requirements.
- **Make-to-order (MTO)**. Products are made from standard raw materials to meet customer specifications.
- **Assemble-to-order (ATO)**. Products are assembled to customer specifications from standard sub-assemblies produced in advance on forecasts. It is a special case of MTO.
- **Make-to-stock (MTS)**. Products are made and put into finished goods inventory in anticipation of customer orders.

These four strategies and the place of the CODP within the manufacturing cycle are illustrated in Figure 2.19. Many factors influence the positioning of the CODP and hence the manufacturing strategy. Companies can use different strategies for different products. The choice is obvious in simple cases. For spacecraft, you are likely to follow an ETO strategy; for toilet paper, MTS. The meat industry is *mostly* MTS because, once an animal is slaughtered, it must be processed, regardless of whether a customer has ordered tenderloin. In other cases it is more complicated.

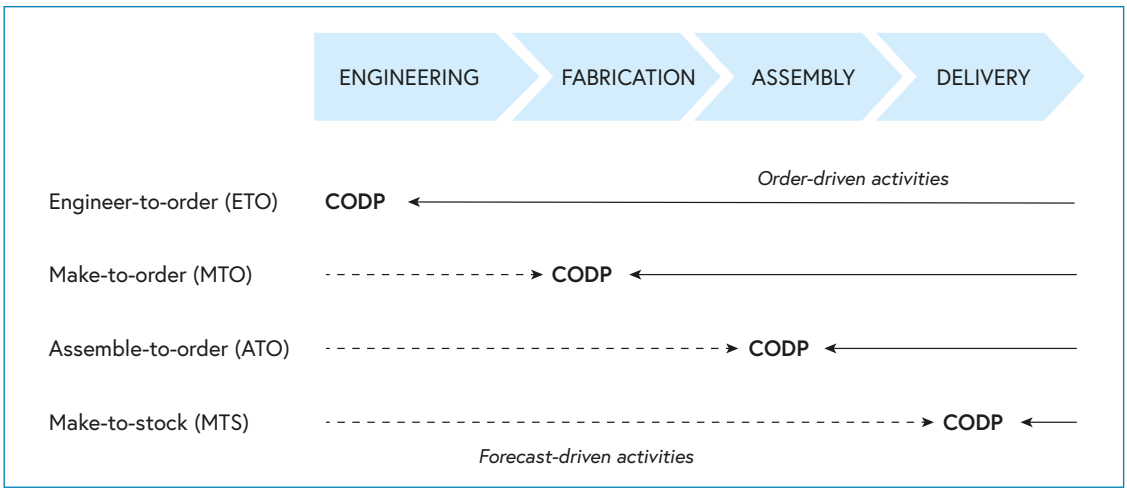


Figure 2.19 CODP and four manufacturing strategies to serve market demand (Credit: Jan Olhager, 2003)[15]

15 Olhager, J. (2003). Strategic positioning of the order penetration point. *International Journal of Production Economics, 85*(3), 319–29.

Cars, for example, are assembled-to-order from standard components (ATO), but these parts are often ordered based on dealer orders, and, as such can be said to follow an MTO strategy. Dealers, however, hold new cars for sale on their lots, that they bought in anticipation of end-user demand, in other words, to stock. While it is MTO for the manufacturer, it is MTS for the supply chain. And, of course, an end-user willing to pay can have a car engineered to order by a manufacturer like Sweden's Koenigsegg.

Moving the CODP upstream, towards raw materials, *postpones* production until the manufacturer is more certain of the demand. Moving the CODP downstream, towards customers, is *speculating* that the products will be sold and investing in making them based on forecasts. Speculation allows manufacturers to plan and use their capacity to a full extent. It may also provide economies of scale. The risk, of course, is that wrong forecasts lead to overages and shortages.

Several factors affect the placement of the CODP. The following are incentives to shift the CODP *upstream* towards raw materials:

- High degree of customization, as at the T-point in Box 2.9.
- Demand that is difficult to forecast.
- Products with expiration dates.
- Products that are expensive to make and store.

The following, on the other hand, are incentives to shift the CODP *downstream* towards the customer:

- High delivery reliability requirements.
- Short-order fulfillment lead times.
- Long production lead times.
- Stable and predictable demand.
- High degrees of product standardization.
- Lack of process capability, requiring systematic overbuild.

Box 2.9 **The T-point**

Many manufactures have one or several specific points in their production processes where the number of product variants goes from few to many. This "variation explosion point" is called a T-point and is usually a good candidate for placing the CODP.

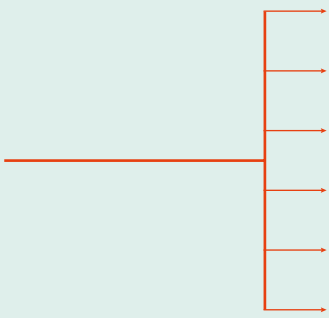


Figure 2.20 The T-point is the place where product variation explodes

Here are some examples of typical T-point locations in different industries:

- Automobile production: Assembly line
- Customized t-shirt shops: Printer press station
- Furniture production: Upholstery
- Luxury yacht production: Drawing board
- Pharmaceutical production: Packaging line
- Computer production: Assembly line

The decision of whether products are engineered or made to order or to stock affects many design characteristics of a factory but is of less importance in managing day-to-day operations. Why? Because all factory operations are in essence make-to-order from *Sales*. Whether Sales is responding to customers or building inventory in anticipation of a promotion campaign makes no difference to *Production*, where it all adds up to the demand they are serving.

2.7. **Focused factories versus capacity factories**

Besides the make-or-buy decision, manufacturers decide what parts to make in which factories and what capacities and competencies each should have. This is also a strategic decision with great implications for productivity and competitiveness.

Wickham Skinner[16] suggested that a good way to organize manufacturing is to make focused factories. According to Skinner, a focused factory is a part of a factory that is dedicated to a product or product family from start to finish. Besides mandating that all the resources needed be available, the concept of a focused factory has no implication on the internal layout of its equipment and workstations.

> **Definition**:
> A focused factory is a part or the whole of a factory with all the human and technical resources needed to make a whole product.

If the plant is organized in focused factories, each area will have its own, dedicated machining, assembly, and test equipment, and each one reporting to a manager with *profit-and-loss* responsibility for the product or product family.

If the plant as a whole is instead organized by function, then there are department managers for machining, assembly, and test, and these departments are usually run as *cost centers*, given an operating budget, and accountable for its use.

Both structures are found among top manufacturers. Toyota's engine plant in West Virginia is structured in focused factories; Honda's in Ohio, by function. Auto parts manufacturers that serve multiple car makers often operate production lines that are dedicated by customer and are, in fact, focused factories. Contract manufacturers, which by definition are capacity factories, on the other hand, commonly organize factories by function.

There are few plants that can be completely split into focused factories, because there usually are operations that the manufacturer doesn't know how to distribute. Machine shops, for example, often include processes like electroplating, heat treatment, and painting that are run as common services.

The majority of focused factories produce in volumes that are high enough to justify dedicated equipment. Occasionally, they are used for products made in low volumes, provided the demand is stable and the manufacturer can find appropriate equipment for this volume. It may mean buying small machines if available, retrofitting old machines, or building machines from scratch.

16 Skinner, W. (1974). The focused factory. *Harvard Business Review, Vol. 52*(3), 113–21.

Summary

The most essential points covered in this chapter were as follows:

1. Manufacturing is more than the acting part of an organization. It is a strategic activity with considerable implications for performance.
2. A company chooses its competitive priorities and seek to build corresponding competitive capabilities.
3. Developing a manufacturing strategy requires balancing an external market perspective – what the market wants – and an internal resource perspective – what we can do.
4. Strategy planning can be assisted with tools such as PESTEL, Porter's five forces, co-opetition, clockspeed, and SWOT analysis.
5. There exists trade-offs in manufacturing capabilities, but some firms have been able to build operational excellence by building capabilities cumulatively.
6. Transaction cost theory can help inform the make-or-buy decision; one of the most strategic decisions manufacturers must make.
7. The customer order decoupling point (CODP) differentiates order-driven activities from forecast-driven activities.
8. Four different manufacturing strategies are engineer-to-order (ETO), make-to-order (MTO), assemble-to-order (ATO), and make-to-stock (MTS).
9. The T-point represents the variant explosion point and is usually a good candidate point to place the CODP.
10. A focused factory is a factory, or part of a factory, dedicated to one or a few product groups.

Test yourself

1. How is manufacturing a strategic activity?
2. Use transaction cost theory to consider whether a manufacturer of wooden furniture should make or buy the following goods and services. Make assumptions when necessary:
   - Screws and nails for the furniture
   - Wooden legs for chairs and tables
   - Catering for employees
   - Assessment of local tax schemes for a new factory site in a foreign country
   - Maintenance of machines
3. What competitive capability is at the base of the sandcone model? Why?
4. What does the smiley curve propose?
5. What is transaction cost theory and how does it inform the make-or-buy decision?

6 What separates the four generic manufacturing strategies conceptually?

7 Consider the following products. What manufacturing strategy are they mostly made with? Where is the CODP?

- Cruise ships
- Printed novels
- Tables sold at furniture retail shops
- Dental prosthesis
- Pizza in a restaurant
- Eye glasses

8 What is meant by speculation and postponement?

9 What product characteristics encourage shifting the CODP downstream? Why?

10 What is the difference between a focused factory and a capacity factory?

Further reading

Beckman, S. L. & Rosenfield, D. B. (2008). *Operations strategy – competing in the 21st century*. McGraw Hill.

Boyer, K., Swink, M., & Rosenzweig, E. (2005). Operations strategy research in the POMS journal. *Production and Operations Management, 14*(4), 442–449.

Cagliano, R., Acur, N., & Boer, H. (2005). Patterns of change in manufacturing strategy configurations, *International Journal of Operations & Production Management*, *25*(7), 701–718

Gray, J. V., Helper, S., & Osborn, B. (2020). Value first, cost later: Total value contribution as a new approach to sourcing decisions. *Journal of Operations Management*, *66*(6), 735–750.

Hill, T. (2017). *Manufacturing strategy: The strategic management of the manufacturing function*. Macmillan International Higher Education.

Miltenburg, J. (2005). *Manufacturing strategy: How to formulate and implement a winning plan*. Productivity Press.

Schroeder, R. G., Bates, K. A., & Junttila, M. A. (2002). A resource-based view of manufacturing strategy and the relationship to manufacturing performance. *Strategic Management Journal*, 23(2), 105–117.

Skinner, W. (2007). Manufacturing strategy: The story of its evolution. *Journal of Operations Management, 25*(2), 328–335.

Slack, N. & Lewis, M. (2015). *Operations strategy* (4th ed.). Pearson Education Limited.

Voss, C. A. (1995). Alternative paradigms for manufacturing strategy. *International Journal of Operations & Production Management, 15*(4), 5–16.

Voss, C. A. (Ed.). (1992). *Manufacturing strategy: Process and content*. Chapman & Hall.

Wheelwright, S. C. (1984). Manufacturing strategy: Defining the missing link. *Strategic Management Journal, 5*(1), 77–91.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

3 Serving Market Demand

The effectiveness of a factory is determined by its ability to serve a market demand. As manufacturers must make decisions about the future, they need to predict it. They may have product planners and marketers with vivid imaginations. Yet, to avoid making products that nobody wants to buy, they rely on established forecasting methods. This chapter presents forecasting as a rational method for predicting customer demand. Although forecasts are needed, they are usually wrong. To make them as good as possible, a mix of quantitative and qualitative forecasting methods are often called for.

When the forecast about the market demand and its characteristics are known, the next step is to ensure the factory capacity needed to serve it. For this purpose, the concept of takt time is particularly useful as a basis for designing and operating many types of production systems. With few exceptions, every factory has a takt time. Think of takt time as the pulse of the factory. Some have a fast pulse, others slow. Some factories have a regular beat, others many variations. The pulse can be the same for all products or vary across products. The concept of takt time helps you pursue the goals of effectiveness and efficiency. The way to increase efficiency is to seek a stable and fast pulse, moving towards continuous processing, but not at the expense of meeting the demand.

Learning objectives

After working through this chapter, you will be able to do the following:

1. Characterize demand patterns for a product by detecting managed variation, trends, cycles, seasonality, and residual variation.
2. Make a simple forecast for a product.
3. Make use of qualitative forecasting methods.
4. Plan the capacity of a new factory.
5. Conduct a Sales & Operations Planning meeting.
6. Calculate the takt time for a product.
7. Use the concept of takt time to improve the efficiency of a factory.

DOI:10.4324/9781351110310-4

3.1. **The art of forecasting**

Anticipating future demand is not easy, and even the best make mistakes. Key business decisions rely on our expectations for the future, so we have no choice but to attempt to forecast it for our own needs. We must also assist suppliers, who depend on our input for their own forecasts.

Forecasts are statements about the future. Sometimes forecasts are so inaccurate that people propose to do away with them and base decisions solely on orders received. Considering, however, that 9 to 12 months may elapse from when management buys a machine to when it is up and running on the shop floor, this kind of decision cannot be based exclusively on firm orders already received. Forecasting is necessary, and no manufacturing organization can avoid the effort of doing it as well as possible.

Forecasts are used in many different ways:

- Long-term forecasts support decisions to build new lines or factories, reallocate products between factories and, sometimes, to shut down factories.
- Medium-term forecasts are used in Sales & Operations Planning (S&OP), which is a negotiation among managers in charge of sales, production, and finance. It supports decisions about the shift schedule of the factory, planning for overtime or employee vacations, and giving a heads-up to suppliers and customers.
- Short-term forecasts are often combined by the Sales department with firm orders into Master Production Schedules that are used to drive production on a daily basis.

The meaning of long-, medium-, and short-term varies by industry. The commercial aircraft industry works to firm orders known 18 months in advance, and long-term means 10 to 15 years. In automotive, long-term can mean 3 to 5 years and medium term can be 3 to 6 months, as exemplified by Toyota adjusting the pace of its production lines every four months. In the toy industry, 1 year is long-term.



Figure 3.1 Forecasting uses existing data and estimated data to predict future events such as sales

Many companies' idea of forecasting demand is to apply an algorithm to extrapolate from the recent past. However, producing a forecast should not be reduced to executing a protocol. This is an area where human decision is called for to take into account not only past sales, but also quantitative and qualitative inputs from the sales force, customers, trade publications, and leading economic indicators.

3.1.1. Principles of demand forecasting

Forecasts are never right. Yet forecasting is an important activity for manufacturing. You need forecasts to plan capacity and production. Here are six general principles for forecasting:

1. Forecasts are always wrong.
2. Forecasts are more accurate for families of products.
3. Forecasts are more accurate for shorter horizons.
4. Good forecasts include an error estimate.
5. Forecasts are no substitute for actual demand.
6. The effect of human behavior should not be underestimated.

Forecasts are always wrong. There is of course a statistical chance that you hit the exact number of sales in a future month, but even that would not be because your forecast and its assumptions were completely right. A good forecaster does not aim to be perfectly right, but to be less wrong.

Forecasts are more accurate for families of products than for single products. This is because variation is spread on more products. The low forecasts balance out the high forecasts, at least in sizable groups. For example, if your factory is manufacturing and selling lemonade with different artificial flavors, your forecast for the family of lemonade products will on average be more accurate than your forecast for strawberry lemonade.

Forecasts are more accurate for shorter horizons. The nearer in the future you forecast, the better the forecast should be. If you were forecasting the demand for lemonade next week, it would normally be a better forecast than the forecast for lemonade eight weeks from now. This is because you have more recent information on past sales, which you can extrapolate from.

Good forecasts include an error estimate. Giving just one number for a forecast is risky. There are two important issues here. First, what band width of possible outcomes are we forecasting? This number can be given by providing an expected standard deviation for the estimate. Second, is there a higher or lower chance for hitting below or above the forecast? This can be represented by a positive or negative skew. If your forecast is conservative or optimistic means a lot for decisions. A simple way of providing error estimates is to provide the forecast with a plus/minus figure. For example, next months, we forecast to sell 20,000 ±3,000 liters of lemonade. Skewness can be included as in, for example, 20,000 (+4,000, -1,000) liters.

Forecasts are no substitute for actual demand. When you receive a number of new orders, you should, if possible, update your forecast. It will then become more accurate. When you know the total actual demand for a period, the forecast is obsolete. But this is the time to calculate the goodness of the forecast and learn from it. Use the forecast error to calculate the forecast accuracy, which is the ratio of the residual to the actual demand, stated as a percentage.

The effect of human behavior should not be underestimated. Humans are making forecasting decisions with bias and judgement. A biased forecast is one that is consistently wrong in the same direction. Many forecasts are biased to please an audience. Studying weather reports in cities, Nate Silver noticed that they exaggerate the probability of rain every day. Nobody complains if it does not rain on a day the weather report said it would. On the other hand, they do complain if it rains on a day the forecast said it would not.

3.1.2. Getting access to market information

Sales forecasting is not just a data analysis problem. Before you can analyze data, you need to acquire it, and many manufacturers receive it too late and at too high a level of aggregation.

Outside of manufacturing, a transportation service company like Uber has records of all its drivers' transactions in real time, and access to other information affecting demand, such as weather, traffic, planned events like sports games or concerts, and unplanned events like earthquakes, floods, or fires. As of 2020, this adds up to over 100 petabyte of data, and a force of 350 data scientists to draw conclusions from this data set that the company can use. No manufacturing company comes close to this amount of relevant data. If you are selling directly to consumers, you have direct access to the end-user of your products. Few manufacturers, however, do this. Most manufacturers rely instead on a distribution network that delays and restricts their access to market information.

Manufacturers decide which products to make, how to promote them, what prices to charge, and how they are going to be sold by others. Apple operates its own stores and Dell sells directly to end-users online. Almost all other manufacturers of consumer goods sell to distributors or dealers, a structure that led to the emergence of marketing as a distinct discipline from sales.

This structure insulates manufacturers from the market. GM's leader in the 1920s, Alfred P. Sloan, introduced the dealer system in the car industry for the purpose of absorbing market fluctuations and allowing factories to produce at a steady pace. As Sloan recounted in his memoirs,[1] in 1924, the market for new cars collapsed but GM did not notice until all its dealers' lots were full.

More generally, in multi-echelon supply chains, each node usually communicates only with its direct suppliers and customers. The supply chain management literature says that you should know your suppliers' suppliers and your customers' customers. In practice, you rarely do, and this myopia, if not addressed, amplifies the fluctuations in volume as you go up the chain, a phenomenon Hau Lee called the "bullwhip effect",[2] as covered in Chapter 14 Supply Chain Management.

By contrast, in its early days, Toyota, in Japan, used door-to-door sales reps, a system that was eventually abandoned as too expensive. Decades later, in 2000, Toyota CEO Akio Toyoda set up a website called gazoo.com for the Japanese home market that provides news about cars and car shows, a marketplace of used cars, tales of road trips, recommendations of games to keep children entertained during rides, etc. By their behavior on the site, the visitors to gazoo.com give Toyota the kind of access to market information they used to have through door-to-door sales. Access to market intelligence is key to achieve Toyota's goal of "One time customer, lifetime customer."

There are other, more quantitative ways for a manufacturer to keep their fingers on the pulse of the market. Barilla, a pasta manufacturer in Italy, worked with its distributors to establish *vendor-managed inventory* (VMI), a system that allowed them to see in real time which products moved and which did not.[3]

1 Sloan, A. (1990). *My years with General Motors*. Currency.

2 Lee, H. (1997). The bullwhip effect in supply chains. *MIT Sloan Management Review, 38*(3), 93-102.

3 Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E. & Shankar, R. (2008). *Designing and managing the supply chain: Concepts, strategies and case studies*. Tata McGraw-Hill Education. Chapter 4, pp. 67-82.

3.1.3. From market demand to materials requirements

In manufacturing, demand is separated into two types:

1. **Market demand**, or the demand for finished products.
2. **Materials requirements**, or the demand for materials, components, or subassemblies that go into finished product is dependent.

Market demand is sometimes called "independent," versus materials requirements, called "dependent." We avoid this terminology because it conflicts with the concept of independent random variables, which is used in demand forecasting.

The important issue here is that materials and components do not need their own forecasting process. Their forecasted demand is simply calculated from the demand for the finished products they go into. For this, production planners use bills of materials (BOMs) for discrete products or recipes for bulk products, as covered in Chapter 10 Master Data Management.

Figure 3.2 shows the difference between market demand and materials requirements for an IKEA Lack shelf. Once the forecast for the finished product is ready, the forecast demand for the wall mount, wooden shelf, screws, and packaging material can be calculated.

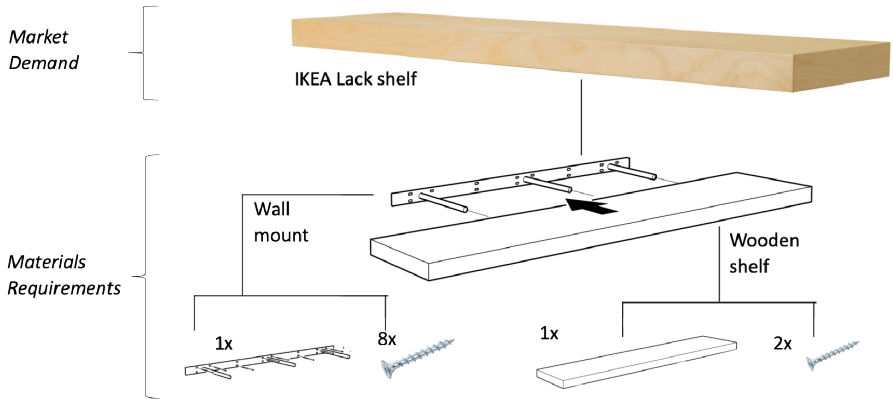


Figure 3.2 Independent and dependent demand of an IKEA Lack shelf

The market demand comes directly from the external environment, while materials requirements are calculated from the market demand. Note that the market demands for different finished products may not be independent random variables. For example, there is likely to be a positive correlation between sales of IKEA Lack shelves and Lack tables. Likewise, demands for bicycles and bicycle helmets are correlated.

> **Stop and Think 3.1**
> Your factory makes teddy bears. What is the meaning of dependent and independent demand for this product? How can you know how many teddy bears to produce? How would you plan how many eye stickers to stock up?

3.2. Demand volume, variety, and variability

To characterize the demand for a factory, it is useful to think about these three "Vs":

- Volume
- Variety
- Variability

3.2.1. Volume

Volume is expressed in pieces for discrete manufacturing or in bulk such as pounds and gallons or kilograms and liters for process industries. For some raw materials, bulk measures are also used in discrete manufacturing as in tons of steel to produce cranes. To achieve a number needed for planning capacity, volumes are aggregated from all the markets served by the factory. Aggregate volumes are easier to forecast accurately than volumes for individual products. The total number of vehicles that will be sold in a specific market this year is generally well known but of limited value for a maker's specific models.

Factories are designed for specific target volumes, and no discussion of layout, tooling, or equipment selection should take place until this volume has been decided. Note that knowing the market size is not the same as knowing the production capacity; deciding how much of the market demand to serve is a strategic decision of the manufacturer.

Volume calculations differ in discrete versus process manufacturing:

- **Discrete manufacturing** is about assembling or disassembling discrete pieces of material using bills of materials (BOMs). Examples of discrete manufacturing are automobiles, mobile phones, computers, white goods, furniture, apparel, etc.
- **Process manufacturing** blends and compounds materials using formulations or recipes. Examples include rocket fuel, shampoo, paint, metal smelting, fermented beverages, etc.



Figure 3.3 Soda production changes from process manufacturing to discrete manufacturing in the bottling process

In process manufacturing, the bulk output is packaged into discrete containers – barrels, bags, or bottles – or become part of solid product units, like steel structures. The point where this happens is called the "*discretization point*."

After the discretization point, the rest of the process is discrete manufacturing, with customization operations like labeling or the inclusion of promotional gifts, followed by aggregation into larger containers like display cases, cartons, and palettes for shipment.

In fast-moving consumer goods industries (FMCG), such as soda production illustrated in Figure 3.4, the discrete operations on packages are often done at 1 unit/second or faster, and cannot be performed manually. Historically, these were among the first operations to be automated. The role of people is limited to

keeping machines supplied with packaging materials, responding to micro-stoppages and breakdowns, and changing over between products.

In discrete manufacturing, volume can range from one unit for a unique one-of-a-kind product to millions of units for standardized components. In process industries, volumes can range from micrograms of special raw materials to tons of steel. In aggregating volumes, additive measures are needed for capacity planning, and multiple units of measure may be necessary for different purposes. For example, in transportation, the amount of cubic feet and the weight of the product determine feasible truck loads, and, in production, it may be the number of components, the amount of labor, or the time required on a bottleneck machine.

> **Stop and Think 3.2**
> Imagine you are asked to forecast the next year's demand for Coca-Cola in your country and hand it over to the bottling plant. What units would you use? How would your forecast look like?

3.2.2. Variety

Variety, or mix, is the breakdown of the aggregate volume among products, and is not easily reduced to a single number and is most effectively represented by a Pareto chart of product volumes, known as *Product-Quantity (P-Q) analysis,* covered in Chapter 7. There is variety between different products, between different models of the same product, and between different variants of the same model.



Figure 3.4 Customization of sandwich at Subway

Customization is the process of adding customer-ordered features to a product. Customization is usually offered as a choice between different predefined features, as when ordering a sandwich at a Subway restaurant.

Full customization can be found in industries that engineer and design products based on each customer's order. Note again that managers must make a strategic decision on how much variety to offer and whether customers will have the opportunity to customize.

3.2.3. Variability

Demand varies over time for many reasons. There is always variability, but it can vary greatly from product to product. Variability can be understood by separating it into five different components:

1. **Managed variation**: Variability stemming from use of sales tactics to influence demand.
2. **Trends**: The general direction of a market. Trends are patterns of growth, decline, or flattening out.
3. **Cycles**: Repeating patterns of growth or decline in sales. Unlike the calendar year, the cycles vary in length, and their timing is at least as much of an issue as their amplitude. For example, the sales of automobiles is following cycles of economic growth and decline. Sales of oil tankers follows

volatility in oil prices.

4 **Seasonality**: Shifting demand patterns across time periods. This can be seasons – spring, summer, fall, winter – particular days of the week or month, annual occasions like Christmas, Easter, Chinese New Year, Ramadan, Mother's day, Halloween, or recurring events like the Olympics or award ceremonies.

5 **Residual variation**: Variability that cannot easily be ascribed to any of the other four types. It seems to follow no pattern or repetition. It can be prescribed to luck or bad luck. For example, if the local football team wins the national league, the demand for beer can soar just on that particular weekend. Natural disasters can stimulate the demand for goods used in reconstruction in the affected area, while hurting the demand for goods used in normal life and parts consumed by affected factories.

Figure 3.5 shows an example of a demand pattern during three recent years for a real factory producing chemical products in Indonesia. The graph shows the sales volumes and the following components of variability: managed variation, trends, seasonality, and residual variation. Note that this is an aggregated demand for all products produced, so variety is not shown.

From Figure 3.5, it is clear that there is a low demand around June and July each year. Considering the three years shown, this looks like seasonal variations. In this case, however, it relates to lower demand in the month with Ramadan, when the large Muslim population of Indonesia has two weeks off. Ramadan, however, is not a seasonal but a religious event, based on a lunar calendar, and it moves back by about 10 days each year. This would be visible if we had plotted 10 years instead of 3.

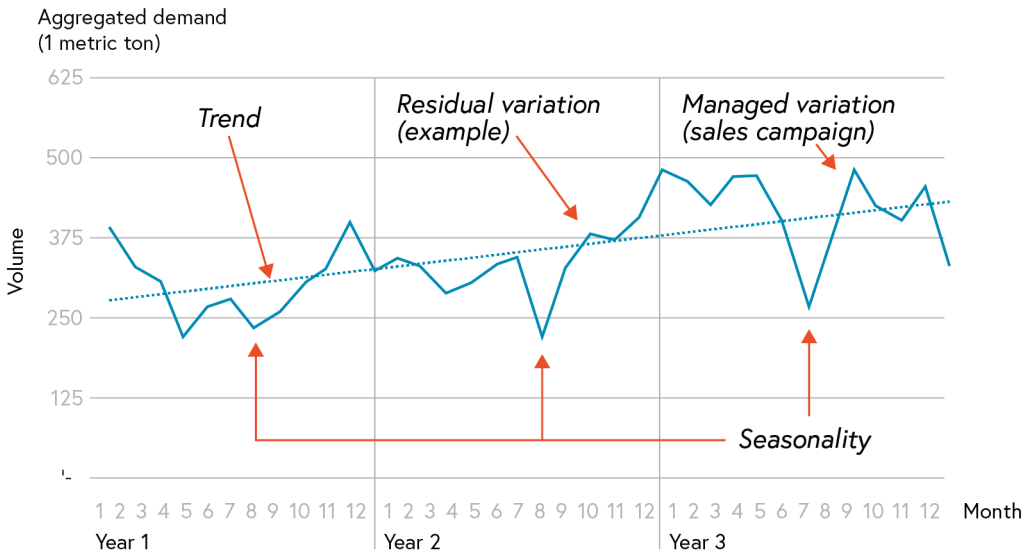


Figure 3.5 The demand of a chemical factory is characterized by volumes, trends, seasonality, managed variation, and residual variations

The arrow pointing to a demand peak in Year 2 looks like residual variation but could also be due to one large incoming order or due to a sales campaign. The straight line in the figure is a trend line drawn for the whole period. It is clear that the trend line can look very different if drawn for a different period. For example, if we consider the trend for Year 1 in isolation, we get a negative trend.

Cycles can be hard to detect just by studying the three years of demand in Figure 3.5, but with a longer time horizon, demand cycles would appear. Cycles in the paint industry, for example, can be related to the start and end of large construction projects, which again can relate to a country's development in gross domestic product (GDP).

An example of cycles is shown in Figure 3.6, which depicts the historical deliveries of Boeing 737, the world's best-selling aircraft in aviation history, compared with its main rival, the Airbus A320. Through about 2010, the sales of both follow cycles of about three to five years. From 2010 to 2019, however, the cyclical pattern seems to be broken and the chart should be a prompt to research why. It is obvious that these patterns are important. Each data point represents billions of dollars of investments by hundreds of airlines worldwide. There is no way a doubling or halving of aggregate demand in 3 years can be due only to fluctuation.

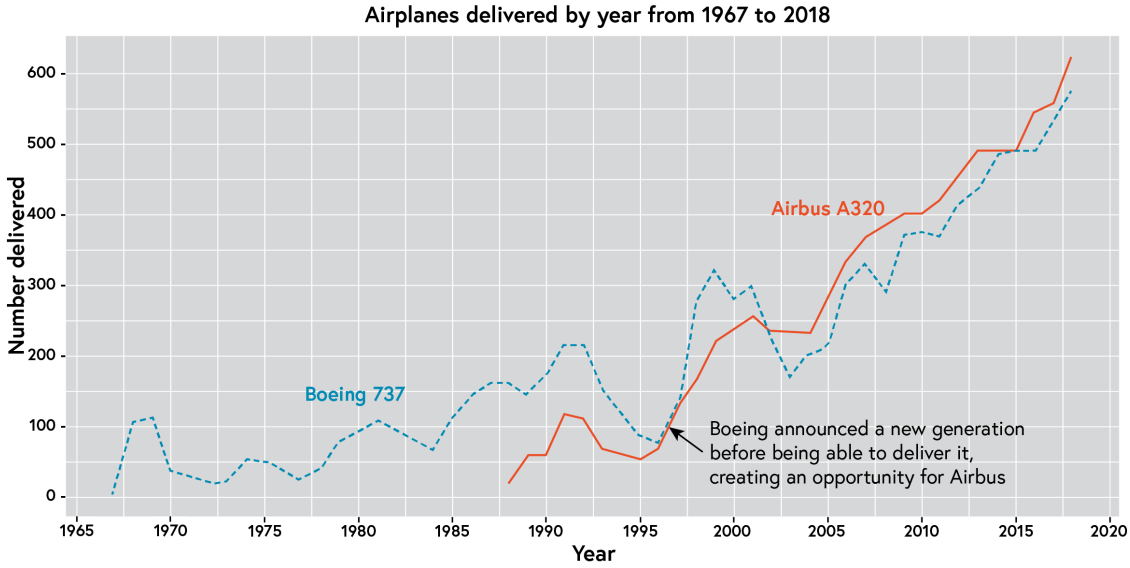


Figure 3.6 Deliveries of Boeing 737s since 1967 and A320s since 1988

The cycles through 2010 are related to the general economic growth of countries where the buying aviation customers are located and the multi-year delivery lead times. Airlines overbuy planes based on optimistic forecasts. Three years later, as the expected growth has failed to materialize, they have a surplus of planes and stop buying. Another three years later, activity has picked up, the airlines are short of planes, and overbuy again...

Why has this pattern been broken? The business has changed on both the supplier and the customer sides. By improving manufacturing operations, the aircraft makers have reduced their order fulfillment lead times to 10 months, making it easier for airlines to forecast. In addition, the airlines have been applying data science to improve their short-term ability to fill planes with passengers and their forecasting.

It was a mistake in 2010 to assume that the 3-5-year cycles would continue. It would have been a mistake in 2019 to assume that the linear growth of the past 9 years would keep going forever. In hindsight, we know that the COVID-19 pandemic changed the demand patterns for aircraft drastically. The chart is useful but insufficient as a basis for long-term capacity planning.

Due to the five components of demand variability, the way production facilities are set up will not be the same for the following products:

- Electricity meters in a mature market, where the demand is for replacement of worn-out units, is constant all year long, and varies little from year to year.
- Toys, where most of the production is sold in the weeks before Christmas and the product goes out of fashion shortly thereafter.
- Food processing, where variations in volume are determined by crop seasons and crop yields, and seasonal variations are dictated by the supply of raw materials rather than by demand.
- Prestige cosmetics, where sales peak around Mother's Day and Christmas.
- Semiconductor production equipment, where the demand is subject to wild swings induced by those of its customers' cyclical industry.
- Mobile phones, or other electronic devices for which there is pent-up demand, with sharp growth until the market saturates, at which point it drops sharply.
- Medicines for epidemic outbreaks such as for example Tamiflu for the bird flu or COVID-19 vaccines for the coronavirus pandemic, where orders can suddenly be gigantic.

> **Stop and Think 3.3**
> You are the project manager for the development of a new production line and its supply chain for a new model of an aircraft competing with the Boeing 737. How does Figure 3.6 affect your decisions?

3.2.4. Managing demand

All components of demand variation, except managed variation, are normally outside the control of the company if it operates in a non-monopolistic market. Managed variation therefore deserves special attention. Companies can "manage" their demand by manipulating it. It can be strategically used to amplify or reduce the effects of the other types of variation. These activities are done by Sales and Marketing, but the Production department must be involved in their planning – or as a minimum be informed. Some methods of demand management include:

1. Long-term contracts
2. Price changes
3. Advertising
4. Incentives
5. Promotional giveaways, like gift-with-purchase or purchase-with-purchase items

Products must be sold. That means that someone needs to do sales and marketing. Although the type of people who choose sales jobs often have different personal characteristics than the one that choose to be in manufacturing, it is essential to create a good dialogue between sales and operations. Sales is the most important way to manage demand, and operations the most important way to fulfill it. The best thing sales can do for operations is to land long-lasting contracts to serve a customer with large volumes of standard products at a price substantially larger than production cost.

A classic way of managing demand is to vary the price of the product. Economists call it "price elasticity," which explains how markets react to changes in price. A risk when making frequent changes in prices is that the customer gets used to it. They refuse to pay full price and wait for the product to come

"on sale" when they stock up on the product. The effect can be large variations in demand that in the end will hurt the efficiency of the manufacturing operations.

In fast-moving consumer goods (FMCG) industries, promotion is so essential that an established brand dies in 6 months in its absence. For example, Coca-Cola leverages their marketing capabilities to increase sales around large sporting events, such as the Olympic Games and World cups (Figure 3.7) and holiday seasons. Sales campaigns such as "buy two, get one more" or large discounts are also very effective in increasing demand for many products. Campaigns are less common in business-to-business environments.

Incentives are another way to manage demand. An everyday example is how airlines and supermarkets use loyalty programs to make customers return. Similar strategies can be used in sales of manufactured products. For example, one paint manufacturer dropped lottery coins in their paint pots to encourage painters to buy their pots instead of competitors. The example in Box 3.1 shows how such efforts can misfire.



Figure 3.7 Marketing of soft drinks during FIFA 2018 Soccer World Championship

To increase demand, the Sales department can also offer additional products or services. For example, a professional tools manufacturer can offer a full range of products to serve whatever need the customer has. The manufacturer can also add services such as financing, free training, efficient spare part logistics, prolonged guarantee, and a range of other incentives. Adding services to a manufactured product is called "servitization."

Box 3.1 **Sales campaigns gone wrong**

The markets for consumer goods are routinely disrupted by new, radically different products, and incumbent suppliers scramble to respond. If you sell detergent powders, it may be liquid detergents; if espresso coffee in 1 lb bags, individual capsules; if breakfast cereal, multiple flavors. In each case, you must quickly develop and promote a response.

One example of a quick response is to launch a sales campaign. For example, "buy-one-get-one-free" is a common promotion scheme. But it must be designed carefully, otherwise it may backfire. A European detergent company once made a sales promotion that did not go exactly as planned. Marketing managers decided to add a gift coupon for another detergent box in the detergent box. The problem was that the second detergent box also included the same gift coupon. Before the company realized, customers had stocked up on several units for the price of one, the production volumes rose sharply but without generating any revenue.

This is a typical example of mistakes people make under stress. To avoid the issue, the detergent company could pack and distribute the gift units separately, which adds complexity to operations for the manufacturer, distributors, and retailers. In the heat of the commercial battle, people sometimes make decisions that make life in operations unnecessary difficult.

3.3. Quantitative forecasting methods

The purpose of demand forecasting is not to establish the presence of periodic variations but to quantify them. In forecasting, the availability of timely and clean data constrains the methods that can be used. New products, by definition, have no history, and you have no choice but to rely initially on qualitative methods. As the history accumulates, you can then migrate to the quantitative methods.

The techniques described in the following sections are not mutually exclusive. To the extent you have records of past sales, you can apply time series analysis to produce a forecast and then use causality models to refine it based on other sources of market data.

3.3.1. Time series analysis

Quantitative forecasting techniques use past values to predict future demand. They use only past data, and see no other variables as important. Since forecasting uses a series of data of evenly spaced numerical data, the right statistical approach is time series analysis. The following time series models are common forecasting techniques, dating back to the 1970s:

1. Naïve approach
2. Trend projection
3. Simple moving average
4. Weighted moving average
5. Exponential moving average
6. Exponential smoothing
7. Autoregression (AR)
8. Autoregressive-moving-average (ARMA or ARIMA)

Greek mathematician Spyros Makridakis has been organizing forecasting competitions since 1982. As measured from 1982 to 2018, in the Makridakis Competition comparing the accuracy of time series forecasting methods, the above methods have outperformed newer ones based on machine learning or neural networks. In the 2018 competition named "M4,"[4] however, an Uber data scientist named Slawek Smyl won with a hybrid method called ES-RNN, for Exponential Smoothing-Recurrent Neural Networks.[5] Uber data scientists report applying this method to anticipate the effects of extreme events like holiday travel, concerts, weather emergencies, or sporting events. As of this writing, we know of no examples of uses of ES-RNN to forecast the demand for manufactured products.

As R&D in forecasting intensifies, the frequency of the Makridakis competitions has increased. M5 held in 2020[6] with teams challenged to forecast the sales by Walmart of 3,049 products in 10 stores for 28 days based on a 5.5-year history of daily sales supplemented by external variables including, for examples, special days, weather events, and promotions. Out of ~20,000 candidates, the winner was LightGBM, "a gradient boosting framework that uses tree based learning algorithms."[7]

4 Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusions and way forward. *International Journal of Forecasting, 34*(4), 802-808.

5 Smyl, S., Rangagathan, J., & Pasqua, A. (2018). *M4 forecasting competition: Introducing a new hybrid ES-cRNN model.* Uber Engineering.

6 Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2021). The M5 competition: Background, organization, and implementation. *International Journal of Forecasting.* In press.

7 LightGBM's documentation. URL: https://lightgbm.readthedocs.io

Note that a general understanding of the methods is usually sufficient for factory managers and engineers. Some of these methods can also be combined. The time when quantitative forecasting was done manually is gone. Modern business software takes care of this automatically, and the task has been reduced to choosing parameters and clicking buttons in the forecasting software. The key is to understand the implications of the choices you make. You must also be able to use qualitative input to update the forecasts produced.

Naïve approach

In the naïve approach, we simply anticipate that the future demand will be the exact same as in the past. This analysis is quick and easy. It is also easy to understand for everyone involved. In a very short term, it is difficult to improve on the naïve approach. For example, if the sun is shining right now, we can predict it will also shine in 10 minutes but perhaps not in 2 hours.

If you ship to customers based on orders received from purchasing agents, you have no visibility into the consumption of your products between orders and naïve forecasting may be ineffective. If, on the other hand, you receive direct orders from the customer multiple times per day or have a vendor-managed inventory agreement, you can use naïve forecasting for the immediate term.

Trend projection

With a demand that is expanding, contracting, or pent-up, we can fit a curve to the data as a function of time. It can be a straight line showing, for example, that sales have increased on the average by $200K per month for the past 18 months. We can then expect this to continue for some time.

Pent-up demand affects innovative products that no one needed before they existed, but no one can do without once they do. Demand initially rises exponentially – with the demand for period n+1 being a fixed fraction of the cumulative sales up to period n – but tapers off and crashes when the market saturates. If the trend model is well built, the residuals around it have a 0 mean but may or may not be autocorrelated. A trend model lets you look beyond just the next period, but how far into the future a trend continues is always an open question.

Moving average models

Simple moving average is the first of several methods to predict the next value from the recent past. The simple moving average model predicts that the next value will be the arithmetic mean of the past n values. It smooths out fluctuations and is easy to calculate but gives the same weight to data from the last period and from n periods ago. If there is an upward trend, it will systematically underestimate future values. Smoothing with arithmetic means is also imperfect in that it sometimes sharpens peaks and valleys instead of eroding them.

The *weighted moving average* method assigns a heavier weighting to more recent data points since they are more relevant than data points in the distant past. A simple way of thinking about weighted moving average is shown in Table 3.1. The assigned weighting should add up to 100% but are otherwise arbitrary. In the presence of a trend, it will also be biased.

Table 3.1 Example of assigning weights to months in the weighted moving average method

Month	Average sales	Assigned weight
January	3,580	10%
February	2,978	20%
March	4,100	30%
April	3,870	40%

> **Stop and Think 3.4**
> Consider Table 3.1 and imagine you are at the last day of April, and want to make a forecast for May. What would be the forecast using the simple moving average method and the weighted moving average method?

Exponential smoothing

Exponential smoothing is a special case of weighted moving average, using a decreasing exponential function to assign the weights. It is a rule-of-thumb technique with no theoretical justification beyond assigning decreasing weights to periods that are more distant. It is used for computational convenience.

Autoregression models

Autoregression is the prediction of the next value of a time series as a linear combination of its past values, with coefficients based on the correlations between past values and calculated so as to minimize the mean square residuals. The coefficients do not have to add up to 100%, and trends do not cause a bias. The range of past values included is determined by the autocorrelation. Today's sales are strongly correlated to yesterday's, but not so strongly to the sales of 10 days ago, and very little to those of 2 months ago.

The residuals of the autoregression model are supposed to be a white noise – that is, Gaussian, centered, and independent. When they are not, the AR estimate can be improved by adding a moving average of past residuals. These models are called *autoregressive-moving-average (ARMA or ARIMA)*. How far back to look in calculating the moving averages can be determined by the Akaike Information Criterion (AIC), an estimator of the relative quality of models for a given set of data.

3.3.2. Selecting the right time series analysis

How do you know which quantitative forecasting method you should use? An established method is to test them on past sales data. Collect past sales data and divide them into an initiation set and an evaluation set. Use the first set to develop forecasts with different models and the second set to evaluate their goodness by comparing the mean forecast error (MFE) and mean absolute deviation (MAD) of each model. MFE is a measure of forecast model bias. MAD indicates the absolute size of the errors. The forecast error for a given forecast is:

$$Forecast\ Error = Actual\ demand - Forecast$$

The Mean Forecast Error (MFE) and Mean Absolute Deviation (MAD) are then given by the following equations for n time periods where we have actual demand and forecast values:

$$Mean\ Forecast\ Error\ (MFE) = \frac{\sum_{i=1}^{n} (e_i)}{n}$$

$$Mean\ Absolute\ Deviation\ (MAD) = \frac{\sum_{i=1}^{n} |e_i|}{n}$$

If a simple forecasting model is producing a forecast that is only marginally worse than more advanced models, choose the simple model, as more people will be able to understand how forecasts are produced.

3.3.3. Effect of external variables

Time series models that are based exclusively on past values are usually supplemented by models of the effect of external variables on the demand. Various regression models fit multivariate curves to data and explain the change in one variable through changes in others. By regression, we explore which predictors affect the demand. The goal is to answer questions like the following:

- Which temperatures relate to peaks in the demand for air conditioners?
- How does the mix of age, gender, and socio-economic background of university students affect sales of beverages on campus?
- How does the selling price influence sales of mobile phones?
- Does collocating beer and diapers on supermarket shelves boost the demand for both?

These insights help us understand the market dynamics better and make better forecasts.

The use of scatter plots and elementary regression is taught in mathematics to middle schoolers all over the world and to shop floor operators participating in quality circles. They should therefore be familiar to manufacturing professionals. In reality, however, they are not. They are rarely found on boards in factories and the promotional materials of most software packages for analytics or business intelligence (BI) are limited to pie charts and stacked bars charts, with a few time series and heat maps. In this context, these concepts need to be explained from scratch.

The word "regression" harks back to the observation by Francis Galton in the late 19th century that the offspring of exceptionally tall parents tended to be still taller than average but less than their parents; in his words, they "regressed towards the mean." Today, regression is commonly understood as the straight line running through a scatter plot, fitted by least squares, as shown in Figure 3.8.

When predicting response Y from predictor X, you use the straight line that minimizes the sum of the squares of the travel distances from the points to the line *parallel to the Y axis*, and, conversely when you are trying to predict X from Y. When you look for the line that minimizes the sum of the squares of the euclidean distance from the points to the line, you don't predict one variable from the other but you find the line that best summarizes the scatter plot, known as regression line.

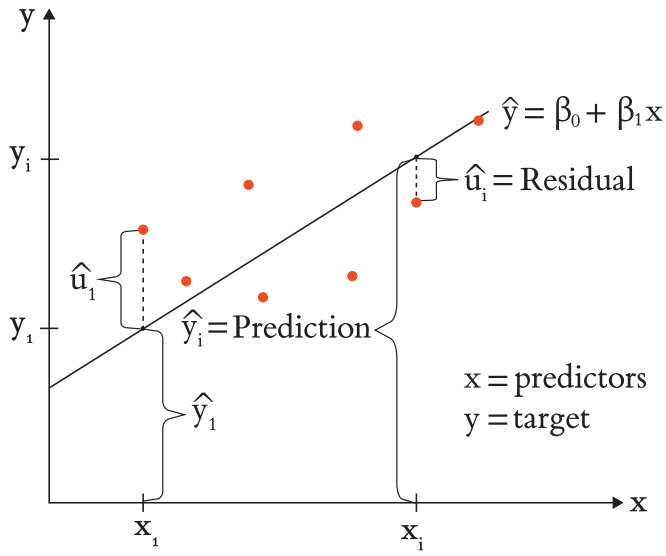


Figure 3.8 Regression analysis

Regression extends naturally to the prediction of a vector, like the demand for all the products in a family, from a vector of predictors like the price of oil, the dollar-to-euro exchange rate, the unemployment rate, etc. The predictors can be categories, like gender or geographical area.

In multiple dimensions, the result comes in the form of formulas with coefficients and cannot be easily visualized as in Figure 3.8. The formulas are not limited to linear combinations of the predictors; they can include terms that are functions of one or more predictors, using, for example, splines to fit curves other than straight lines or hyperplanes.

When there are many categorical predictors, regression *trees* are commonly used, providing answers that are constant for each combination of categories saying, for example, that women aged 18 to 35 are 10% more likely to buy the product than the rest of the population.

Many advanced techniques used outside of manufacturing have yet to be applied in it. The ES-RNN method for Uber referenced in the times series section uses both historical data and external factors. Other techniques Uber is experimenting with include:

- Quantile Forest Regression (QFR)
- Support Vector Regression (SVR)

Even for this simple kind of regression, the software tools of the 2020s go beyond just providing a formula. They also help you evaluate the quality of the model in terms of how well it accounts for the variations in the response variables without overfitting – that is, using so many predictors that the formula models the noise in the set of data you use to generate it – also known as the *training set*. Overfitted models fit their training set closely but are ineffective in predicting new values.

Modern software rank predictors by their ability to account for the variability of the response variable and cross-validate the model by, in turn, removing each data point from the training set, recalculating the model, and assessing its ability to predict the missing point. They also identify for special scrutiny the data points in the training set with the following characteristics:

- **Outliers.** Points where the response variable is outside the mainstream.
- **High influence.** Points whose removal from the training set causes large shifts in the formula.
- **High leverage.** Points with unusual combinations of predictors.

A review of these data points often reveals errors or omissions but sometimes valid exceptions. Their removal from the training set usually improves prediction performance for most points.

Box 3.2 **Forecasting in the era of e-commerce**

Since around the millennium shift, e-commerce has radically changed how goods and services are sold. E-commerce is the activity of selling and buying over the internet. For forecasting, e-commerce has greatly improved how customer sales data can be captured, stored, and used. E-commerce giants like Amazon take advantage of this. For example, when someone is looking at a product on Amazon's web pages, Amazon knows the probability that the person will buy the product. Such statistics can be used to pick, pack, and dispatch the product even before the customer has completed the online check out.

However, e-commerce also introduces challenges to forecasting. Because the internet is enabling an unprecedented possibility for scaling, sales can become more unpredictable than before. For example, when Jack Ma, founder and president of Alibaba visited Thailand in April 2018, he demonstrated the power of e-commerce by selling 80,000 "golden pillow" durians (a Thai fruit, see Figure 3.9) on Alibaba's online supermarket T-mall within one minute. The introduction of the fruit on the Chinese online mall has the potential to completely disrupt the market, making forecasting and capacity planning a much more difficult yet more important activity.



Figure 3.9 Thai fruit Durian sold on local markets versus on Alibaba's T-mall

Another hallmark of our times is the growth of social media and the related power of "influencers" and celebrities to quickly kill or create a surge in demand for certain products worldwide. It is not unusual for the demand of cosmetics products to surge by a factor of 100 just after a celebrity or influencer posts a video or photo of it on a social media channel.

3.4. Qualitative forecasting methods

Qualitative methods should be used to complement quantitative forecasting methods. During introductions of new products, forecasters often have no choice other than relying fully on qualitative methods, simply because there is no available sales data. As the product matures, forecasters will rely more on quantitative input and less on qualitative input. Four common qualitative forecasting methods are:

- Input from sales and customer service
- Market surveys
- Scenario planning
- Delphi methods

While these methods are structured to collect qualitative information about market demand, the most common approach is to just include the intuition and opinions of forecasters. But it is not a practice to be encouraged, as there is a fine line between intuition and confirmation bias. Forecasts based on intuition that turn out to be right make the forecaster look like a genius. These stories are likely to be told and celebrated. The completely failed forecasts, on the other hand, are less likely to be publicized.

3.4.1. Input from sales and customer service

Sales and customer service representatives are in daily contact with their counterparts in customer organizations, which puts them in a position to collect market information through personal communications that is not in the historical data. Customers, for example, may be preparing promotional actions, a natural disaster may have created a demand spike for some products, or they may be merging with another company, which might trigger a review of supplier relations. This is essential information to take into account when forecasting future demand.

More practically, it is common for forecasters to submit the output of quantitative methods to this group to verify that what the data says is consistent with what they hear. From ongoing personal interactions with the reps, the forecasters know that they should modulate their inputs, based on their optimistic or pessimistic temperament. It is not always true but when there is a discrepancy between the quantitative analysis and consensus of the sales reps, there is usually something wrong with the quantitative analysis.

3.4.2. Market surveys

A market survey is the analysis of the market for a particular product and includes an investigation into the customers' intention to buy. It consists of systematically gathering data about people or companies and then analyzing it to better understand what potential customers need and are willing to pay for.

Surveying the market means asking people their opinions about the product and buying intentions. This can be done with questionnaires, qualitative interviews, focus groups, or other methods:

- **Questionnaire surveys** can be distributed to established customer bases or targeting new demographics. They can be conducted as online surveys, paper surveys, phone interviews or face-to-face interviews. The main weakness of surveys is sample bias because they only produce answers from self-selected participants.

- **Qualitative interviews** involve selecting potential customers and asking them questions about preferences, buying behaviors and buying intentions for a specific or potential product.
- **Focus groups** involve selecting a representative sample of 5 to 8 potential customers and letting them discuss the market potential for a product with a trained facilitator. Several focus groups should be arranged.

3.4.3. Scenario planning

A scenario is a narrative forecast that describes a potential future demand. With scenario planning the analyst considers possible market situations that the manufacturer might face. Plausible "scenarios" are created, and then the likelihood of each occurring and the risks involved are analyzed. In scenario planning, you ask "what if" questions as illustrated in Figure 3.10

The demand pattern that might result from each scenario is quantified. A wide range of possible scenarios may be considered, but there is always the risk of missing the scenario that will actually happen. Scenario planning works best when the forecasters have knowledge about similar products in similar markets, referred to as historical analogies. Scenario planning can be enhanced with simulation of data.



Figure 3.10 What if a popular influencer or celebrity reviews our products? What if a war affects our operations and supply chain?

3.4.4. Delphi methods

Delphi methods is a collective term for gathering and analyzing the opinions of a panel of experts in a particular area. One way to apply the Delphi method is to collect individual opinions anonymously via questionnaires, and circulate the results to other members of the group for interpretation and comments. In a final round, the experts are asked to reassess their original opinion in view of those presented by other experts. This process can be repeated until a group consensus is formed. Delphi methods rectify problems of non-anonymous face-to-face confrontation focus groups.

3.5. Capacity planning

The forecast is an estimate of the demand placed by the market on the company, which is not the same as the demand placed by the company on a production factory. If the market demand is cyclical, for example, the company may opt to collaborate with suppliers to develop a surge capacity and supplement in-house capacity during peak times, as discussed below. The challenge is to match capacity to the demand you have decided to serve, without wasting resources, as in Figure 3.11.



Figure 3.11 Balancing capacity with demand is an important task of factory management

Capacity is expressed as a quantity per period and varies by product. Capacity is not a difficult concept but capacity planning can be more or less difficult depending on the data available on the products and the deadline to provide answers. It is most often done in a spreadsheet, where the factory is treated like a network of pipes through each of which the amount of material that can flow is capped by a single number. This is simplistic. A realistic assessment requires deliberations among managers and engineers, including more sophisticated calculations and simulations.

> **Definition**:
> The capacity of a process is the rate at which it can transform inputs into outputs.

3.5.1. General approach

General Managers may ask the Production department whether they can "increase volume by 20%," without any more specifics and with almost no time to answer. If they are operating 3 shifts/day, 7 days a week, and struggling to keep up with current demand, the short answer is "No." It is not what the General Manager wants to hear. The more detailed answer is "not under current conditions." Perhaps productivity can be increased 20% or the lines can be extended, but not by tomorrow. The ramp-up schedule matters. Which products are ramping up also matters. Whether they are new or current products makes a difference because more details are known about current products.

Capacity planning is evaluating resource requirements against availability. When the requirements obviously exceed availability, as in the above example, the negative answer is immediate. If, on the other hand, you are working only two shifts a day, a 20% increase may be feasible, at least in terms of equipment but recruiting and managing an additional shift is neither cheap nor easy or instantaneous.

Before you answer, you need to know which products the increase is for and how much of each resource each unit of these products consumes. If they are all products already in production, data should be available from their bills of materials (BOM) and routings. Their accuracy must be checked but they are available.

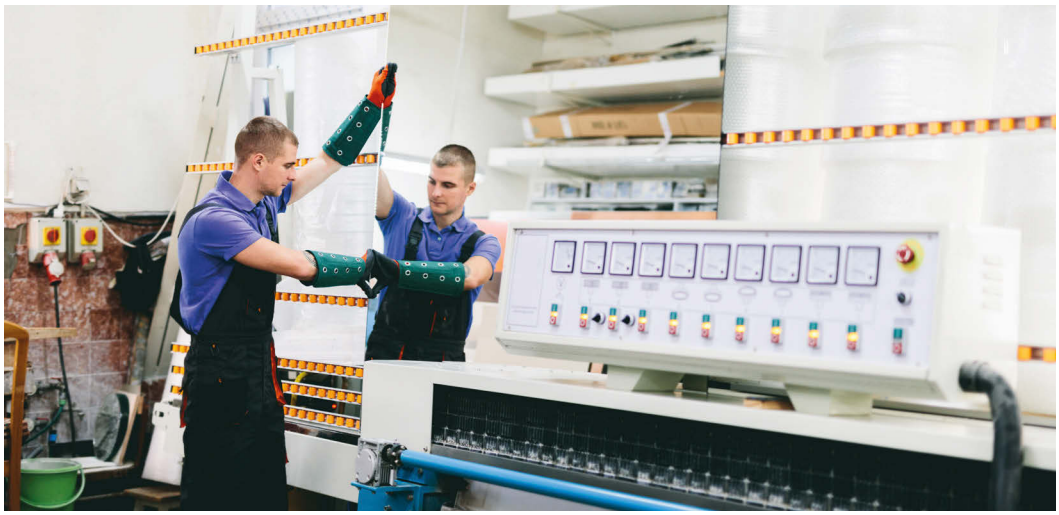


Figure 3.12 Capacity is constrained by the available resources in manufacturing

If, on the other hand, the products are new, you may not have these details and may not want to spend the time and effort needed to acquire them before making the decision to take on the additional volume. The fallback is to *reason by similarity with existing products*.

Informally, this is done by experienced managers who can tell, by looking at the specs of a new product, that, from a manufacturing perspective, it is most like current Product X. It has the same features and tolerances on critical parameters, and is just 20% larger, with 3 holes in a side panel instead of 2, and a new type of paint. The larger product size rules out some of the existing production lines, the extra hole needs a new die, and the new paint process is a question mark that will require experimentation. In the meantime, for every other operation, you work with the available data from Product X.

Formally, this can be done through a model of the influence of product features on processing resource requirements inferred from the BOMs and routings of current products. This can give us a range of requirements in labor, machines, and tooling, for the operations on the new product that it shares with current ones. Until you can refine the numbers through direct observation of operator and machine times in trial runs on the new product, you can add a lower bound of the new products' resource requirements to those of the current products. If the total exceeds 100% of the available time for some machines or stations, we know that the addition of the new products is not feasible under current conditions. If, on the other hand, it is under 100%, you don't know that it is feasible. You need to take a closer look at the busiest resources.

The above calculations are based on the most commonly used model of manufacturing resources in planners' spreadsheet, *the pipe model*, that treats resources like pipes through which the products flow. If product P1 consumes 20% of the net available time of a resource, product P2 another 15% and product P3 another 10%, then, collectively, products P1, P2, and P3 consume 20% + 15% + 10% = 45% of the resource. If you add more products, you simply draw on the slack left by P1, P2, and P3, as shown in Figure 3.13.

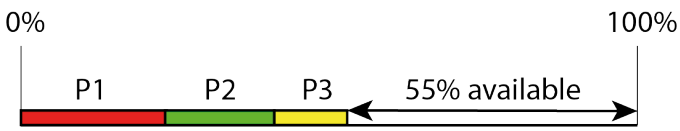


Figure 3.13 Product requirements in the pipe model

In the pipe model, capacity is calculated as follows:

$$Capacity = \frac{Total\ time\ available}{Time\ needed\ for\ task}$$

If a truck factory works on 1 shift for 8 hours a day and every 10 minutes a new truck comes off the line (takt time of 10 minutes), the capacity of the factory:

$$Capacity = \frac{8\ hours/day}{10\ minutes/truck} = \frac{480\ minutes/day}{10\ minutes/truck} = 48\ trucks/day$$

Among its many oversimplifications, the pipe model ignores the saturation effects that are observed at high utilizations when the work is disrupted by random events like microstoppages or late deliveries of materials. In the pipe model, the products flow through the resource in the same time regardless of the utilization u, as long as $u < 100\%$. In a broad range of cases, the time needed for one piece to flow through the resource, including the waiting due to the disruptions, rises like $1/(1\text{-}u)$ as u approaches 100% (Box 3.4). This concept, and the effect of improvements, are illustrated in Figure 3.14.

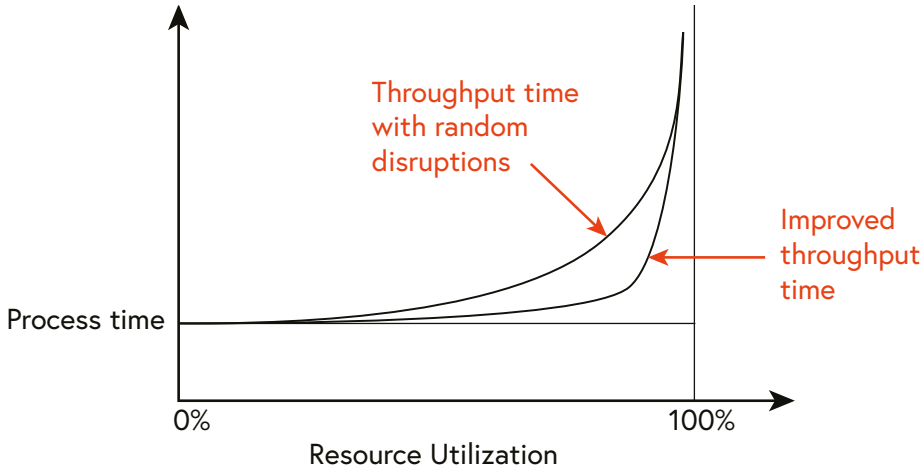


Figure 3.14 Resource saturation

The most common workaround in the pipe model is to avoid loading resources to more than 85% of their nominal capacity, keeping the 15% remaining as an allowance for disruptions. There are two major issues with this approach:

1. The choice of 85% as the threshold is arbitrary.
2. Redefining 85% as the "new 100%" breeds fatalism about process improvement.

Over a long enough period of time, from the executive suite, production resources look like black boxes, with incoming and outgoing flows of products. When you look closer, the only production resources that approximately operate this way are comprised of a large number of identical machines or stations working in parallel. But this is not a common setup. Real manufacturing resources are rarely divisible into many smaller units working in parallel at exactly the same speeds and in the same way. There are usually fewer machines than products working in a multiplicity of ways that impact the throughput time, the labor requirements, and the output quality.

Box 3.3 **Little's law**

In manufacturing, with some exceptions, the benefits of using Queueing Theory do not justify learning its math. The exceptions include Little's law and Kingman's formula. Little's law by John Little[8] is a simple formula for the context shown in Figure 3.15 to estimate the time to turn a blank into a finished good – that is, the lead time in the process.

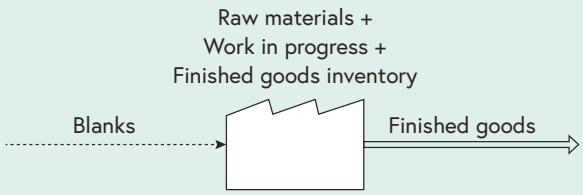


Figure 3.15 Context of Little's Law

In steady state, Little's law relates the mean inventory, lead time, and throughput. It is explained in Figure 3.16, in the simple case of constant throughput and lead time.

$$Inventory = Lead\ time \times Throughput$$

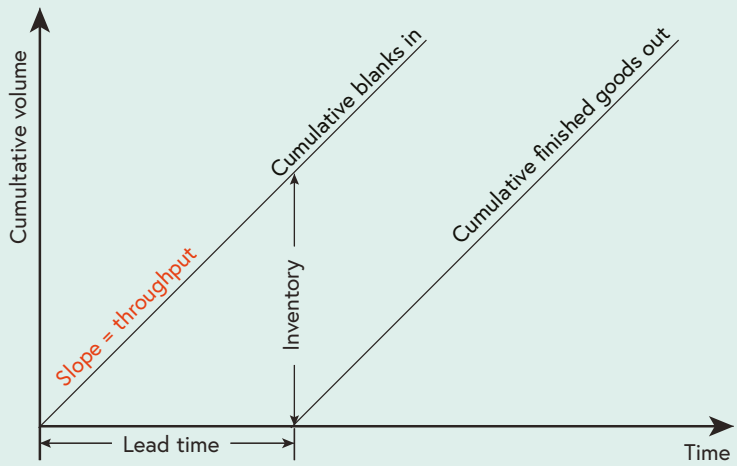


Figure 3.16 Little's law with constant throughput and lead time

The cumulative count of input blanks is a straight line going up over time, and so is the output of finished goods offset by the lead time. The vertical distance between the curves is the inventory. The slope of each curve is the throughput, and the ratio of the inventory to the lead time. Little's law remains valid when both rates of arrivals of blanks and departures of finished goods are allowed to fluctuate randomly around an average.

Since we can count inventory and measure throughput, we can infer average lead times from just this data. One snapshot will not give you an accurate estimate, but it is still considerably easier to take a few snapshots of a production line to get a more accurate estimate than it is to research history. The point is to get close to an answer that would take much longer to get if you actually had to be accurate.

Box 3.4 **Kingman's formula**

Unlike Little's law, Kingman's formula[9] is mostly used qualitatively, as the *explanation of the rise in lead time at capacity saturation* (see curve in Figure 3.14):

$$E(W_q) \approx \left(\frac{u}{1-u}\right)\left(\frac{c_a^2 + c_s^2}{2}\right)\tau$$

It says that, as the utilization u rises towards 100%, the mean waiting time W_q increases like $1/(1-u)$. τ is the mean lead time (also known as "service time" in service processes) of the resource in Figure 3.14; c_a and c_s are respectively the coefficients of variation of the interarrival times and the lead times. The only restrictions on the distributions of arrival and service times is that these parameters should exist.

Kingman's formula is rarely evaluated numerically, because it needs coefficients that are not easy to get, especially on hundreds of resources. Its main value is that it establishes that the saturation pattern in Figure 3.14 is, if not universal, at least very widespread, and that queueing times can be contained by staying just far enough away from $u = 100\%$.

Stop and Think 3.5
Consider the curves in Figure 3.14. If the resource is a road and the process is you driving home, do you think the exponentially increasing process time will occur as the road utilization reaches 100%? What does that teach us about traffic capacity planning?

Effect of batch size

For example, some machines process a single piece at a time; others, a load of identical pieces; others yet, sets of matching pieces… In the pipe model, cumulative volume, through one operation for one product over time, is a straight line but it only truly is if the product is a liquid or a powder pouring into a container. With one-piece flow, it is actually a staircase with such small steps that it looks like a straight line.

If the operation is performed by the *batch*, it looks different depending on whether you are considering it (1) with the range on the time axis that is long with respect to the processing time of a batch or (2) with the range used for planning the work of production operators. Then the steps are high, and the cumulative volume function cannot be confused with a straight line within a shift or a day. See Figure 3.17.

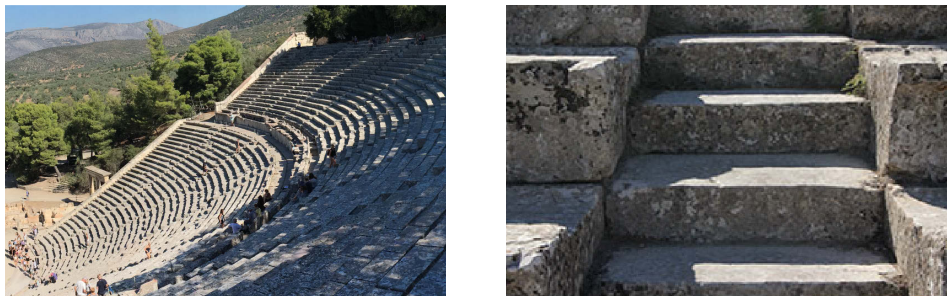
8 Little, J. D. C. (1961). A proof for the queuing formula: L = λW. *Operations Research*, *9*(3), 383–387.

9 Kingman, J. F. C. (1961). The single server queue in heavy traffic. *Mathematical Proceedings of the Cambridge Philosophical Society*, *57*(4), 902.



Figure 3.17 Effect of batch size on output

The corporate planning staff is far enough for the staircase to look like a straight line again. Up close, to production operators, it doesn't, as shown in Figure 3.18. And the lumpiness of the process affects their work, the lead time through the operation, and the quality of the output.



Stairs from afar Stairs up close

Figure 3.18 Staircase volumes as seen from headquarters and from the shop floor

Each piece must wait for a full batch to accumulate before it is processed, and it waits again to be picked up from a processed batch. Between the buffers in front and back of the operation and the batch in process, there are, in steady-state, an average of two batches of work-in-process that need to be managed, and a lead time that is proportional to the batch size.

In addition, the wait delays the discovery of quality problems and this is a strong motivation to pursue one-piece-flow. Finding, or developing, a machine able to process a single piece in 1/50th of the time the current machine takes for a 50-piece batch is sometimes such a technical challenge that the transition to one-piece flow has to wait for new technology. It is often feasible, particularly when the current machine is underutilized and the alternative does not have to match its capacity. None of these issues are perceptible from headquarters.

Effect of mix

If there is only one machine, product P1, for example, occupies 100% of the resource 20% of the time, not 20% of the resource 100% of the time, with time-consuming setups between products, as in Figure 3.19. When there are multiple machines, they are often not identical. You have machines with overlapping capabilities and different process times, load sizes, and setup structures for the different products.

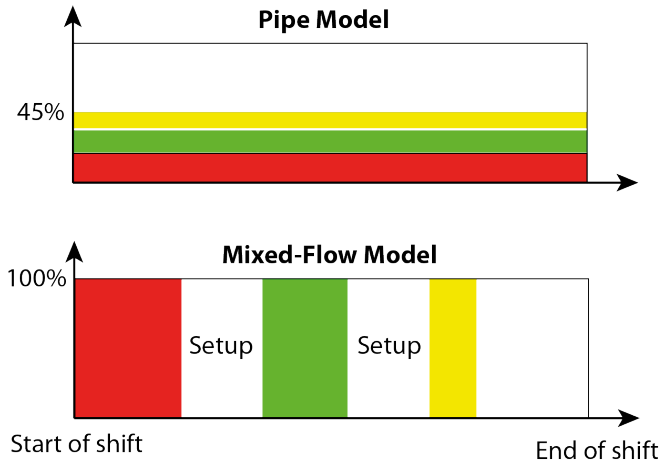


Figure 3.19 Part of the resource all of the time versus all of the resource part of the time

These details determine whether the demand that keeps a particular resource 90% busy according to the pipe model is actually feasible under current conditions. They are observable in actual operations on the shop floor and generally known to the first-line manager and the industrial or manufacturing engineers who support them but some are not known at corporate headquarters and not represented in ERP systems.

Setup times, for example, are generally not a function of just the product you are setting up a machine for but of the pair of products you are changing from and to. It is a matrix of times. On the shop floor, this is basic; in the best-selling ERP product, setup matrices exist only in the "Advanced" planning module.

A capacity planning effort that goes deeper than the pipe model must involve the local organization in charge of production lines and individual operations, not only because they have the data, but also because you want them to own, be responsible for, and committed to the numbers. In addition, the local organization is in charge of the productivity improvements or investments that may be needed to increase capacity. The most effective way to refine the answers provided by the pipes model is to poll the local organizations. They can then provide better numbers based on everything they know.

Effect of variability

Trends, managed variation, seasonal variations, cycles, and residual variation in demand must be taken into consideration in planning manufacturing capacity, either by setting it high enough to meet the peak demand or by setting it lower and building ahead, or by a combination of both approaches. These decisions should be based on a quantitative analysis of the structure of the demand as it varies over time.

It can be focused on one product or on a product family. In the latter case, the aggregation of demand for multiple products may require caution when not all products in a family require the same amounts of resources, and the aggregation of demand may require assigning weights that are based, for example, on the number of components in each product.

On an electronic spreadsheet with sales data, a manager can extract growth trends using built-in analysis tools, filter out daily fluctuations with moving averages, and assess the responsiveness of manufacturing by plotting production and sales data on the same chart. Upstream in the supply chain, component makers see demand fluctuations amplified rather than dampened. The supplier can base capacity planning on the real underlying demand, using finished goods stocks to absorb the artificial fluctuations introduced by the bullwhip effect.

Capacity planners should use at least one year of daily data, but the farther back the data goes, the better. Five years of daily values is well within the processing capacity of commonly available spreadsheet software. If daily data is unavailable, we should use weekly data; if unavailable, monthly data. If all that is available is a stack of invoices, then it is worth going through, because it contains basic business information that management should have.

3.5.2. Types of capacity

For simplicity, it can be useful to separate three types of capacity, as seen in Figure 3.20:

- **Design capacity** is the expected output of an operation when there are no stoppages – the theoretical ideal.
- **Effective capacity** is the capacity we have planned to use, considering planned maintenance, breaks, and time buffers.
- **Actual capacity** is the capacity we have available after subtracting both planned losses and unplanned losses such as breakdowns, sick leaves, or missing parts.

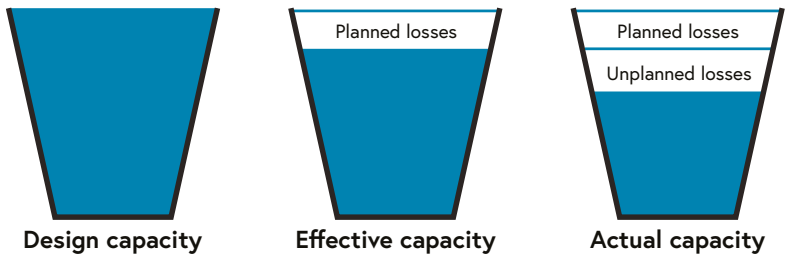


Figure 3.20 Types of capacity

We can use these types of capacities to calculate a couple of useful metrics:

- **Efficiency** is the actual capacity shown as a percentage of the effective capacity of the operation. It demonstrates how well the operation is working to expectations.
- **Capacity utilization** is the actual capacity shown as a percentage of the design capacity of the operation. It demonstrates the degree to which the resources are utilized.

3.5.3. **The capacity of a factory is elastic**

The capacity of a plant and of operations conducted in it is, in fact, elastic. The time required to perform an operation is not a physical constant. It is reduced continuously by improvements in work methods, tools and fixtures, controls, and process programs. And these reductions can be massive, as in, for example, cutting the milling time on an aircraft part from 24 hours to 8 hours in 14 months. Or, in an extreme case, eliminating the process completely by replacing it with a 1-hour 3D printing process overnight. In this case, a capacity plan based on the assumption of a 24-hour process would have been grossly pessimistic.

The amount of work time is also flexible. It is generally not possible to run machines 24 hours per day, 7 days a week, for months without failures but 22 hours per day is feasible. The difference between running a factory with 1 shift per day for 5 days a week versus 3 shifts per day for 7 days a week is a factor of 4. Even if it was constant, human operators' work pace is not. Consider for example the American wind turbine manufacturer in Figure 3.21. There are many reasons why factory capacity is elastic in this factory. Most obviously, the individual workers can adjust their work pace.



Figure 3.21 Factory capacity is elastic: For example, operators' work pace can differ considerably in a wind turbine factory

In addition, plants often have equipment with overlapping capabilities requiring different amounts of time to perform different operations. Machine A may be faster than Machine B on all operations, but 10 times faster on Operation X and 30% faster on Operation Y, giving Machine B a comparative advantage on Operation Y. The allocation of work among such machines also makes a difference.

3.5.4. **Level production versus chase demand**

There are two extremes in production planning; trying to follow every little variation in demand versus trying to smooth out the production. These two polar strategies are called "chasing demand" and "level production." They are illustrated in Figure 3.22.

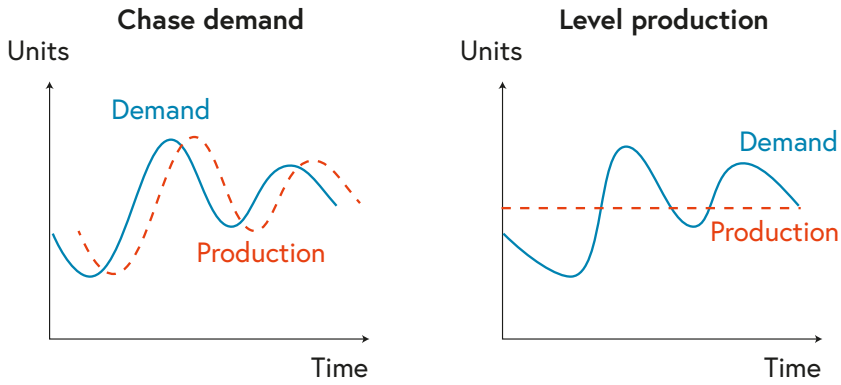


Figure 3.22 Chase demand versus level production

Chasing the demand requires a highly flexible capacity and heavy use of forecasting. To be cost-efficient, the operation should expand and contract its capacity to follow the demand pattern. This can, for example, be done using a temporary workforce. Chasing demand obviously carries the risk of having excess capacity at times and too little capacity at other times. One way to mitigate the risk of not being able to meet demand is to outsource production to contract manufacturers in times of peak demand.

The level production strategy uses inventory to smooth the demand. Demand fluctuations above the production level are served by inventory. In times of demand below the mean, production builds inventory. This form of production is the most resource efficient when seen from a production perspective. It requires, however, that demand is relatively stable and predictable so that the production level can be set and the firm avoids large stock-outs or excessive stock levels. Note that level production does not work for products that cannot be stored, which includes many types of services and highly perishable goods. It also makes no sense to produce highly customized and engineer-to-order products with a level strategy.



Figure 3.23 A Boeing 737 (Credit: Anna Zverova/Flickr)

Figure 3.24 shows how these strategies were both used for the Boeing 737 (see Figure 3.23). The differences between the order and delivery curves tell a story. The smoothness of the delivery curve shows that the company cannot adjust its capacity up or down to follow the order cycles. Also, the milder peaks of the delivery curve lag those of the order curve by two to three years, reflecting the industry's order fulfillment lead times.

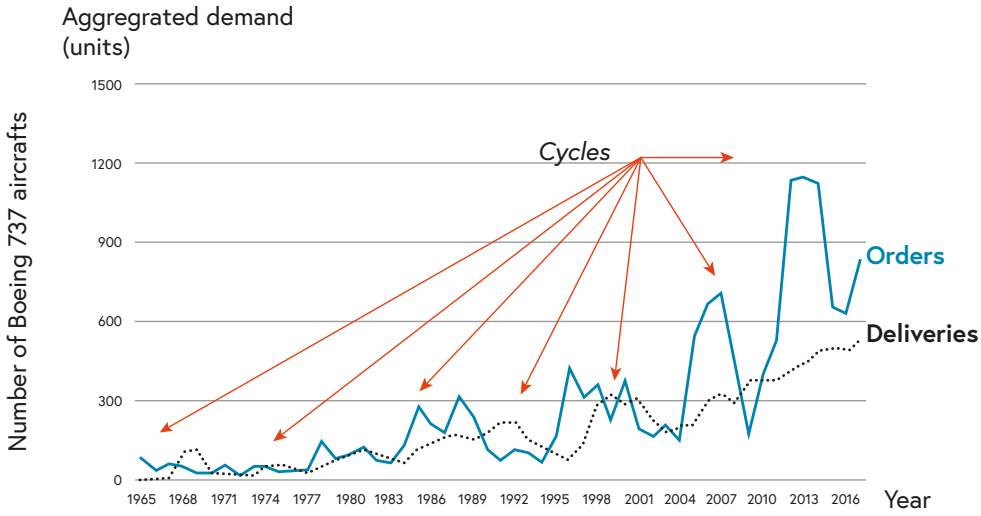


Figure 3.24 How deliveries follow orders for Boeing 737s, 1965–2018

One effective strategy is to decouple processes in a value stream, usually at the variant explosion point, and let parts be produced with level production up to that point and plan for chasing demand after that point. Box 3.5 shows a classic example of strategically selecting the customer order decoupling point.

Box 3.5 **Frying pan fabrication and assembly**

The frying pan company Fry & Friends wondered how to manage an ever-growing inventory of some finished frying pans in their warehouse, while they at the same time always seemed to be short on other pans their customers asked for. A year ago, the marketing team at Fry & Friends had figured out that coloring the handles of the pans in different colors and offering several different sizes of pans could triple the sales of the company. With this insight, the factory had started to stock up on frying fans in different shapes with colored handles, as shown in Figure 3.25.



Figure 3.25 Examples from Fry & Friends' new product offering

After attending a course in manufacturing management, a young quality engineer suggests decoupling the production process in forecast-driven activities versus order-driven activities. The customer-order decoupling point (CODP) was to be placed in a warehouse of semi-finished products. Fabrication of the different types of pan bodies and pan handles should be forecast-driven and planned using a level production strategy, and the assembly of finished pans should be order-driven and follow a chase demand production pattern.

During peak demand before Mother's Day, Christmas, or Black Friday, additional workforce was hired to help assemble frying pans. As a result, Fry & Friends considerably lowered their inventory levels, freed up working capital, and became responsive to shifting demand.

3.5.5. **Sales & Operations Planning (S&OP)**

Medium-term planning considers tactical decisions about what to produce over the next weeks and months. In 1987, consultant Dick Ling introduced the term *Sales & Operations Planning (S&OP)* to describe a process to align and synchronize sales and production capacity.[10] In S&OP, managers from different functions meet and discuss forecasts and capacity on a regular basis, which supports decisions on such issues as overtime or the hiring of temps and provides information to pass on to suppliers for their own planning. S&OP is for product families, not individual products, and for the next three to six months, not for tomorrow and not for the next five years.

S&OP is usually done once a month. The key part of S&OP is meetings, which involves iterations of both number crunching and human judgement. In globally active firms, monthly S&OP meetings cannot be held and are replaced with online conferences. The managers meet to discuss and agree on the next months' deliveries and activities. It reviews the business plan in light of new information about the demand and any changes that may have occurred or need to be made in operations.

The meeting participants represent different functional departments and are empowered to make decisions. S&OP meetings breakdown silos and allow the firm to operate as one entity in the market. Plans are negotiated between the managers of different functions, based on the demand forecast and their understanding of their capacity. The flows of information in S&OP are shown in Figure 3.26, with the output of the process in red. S&OP results in aggregate plans at the product family level.

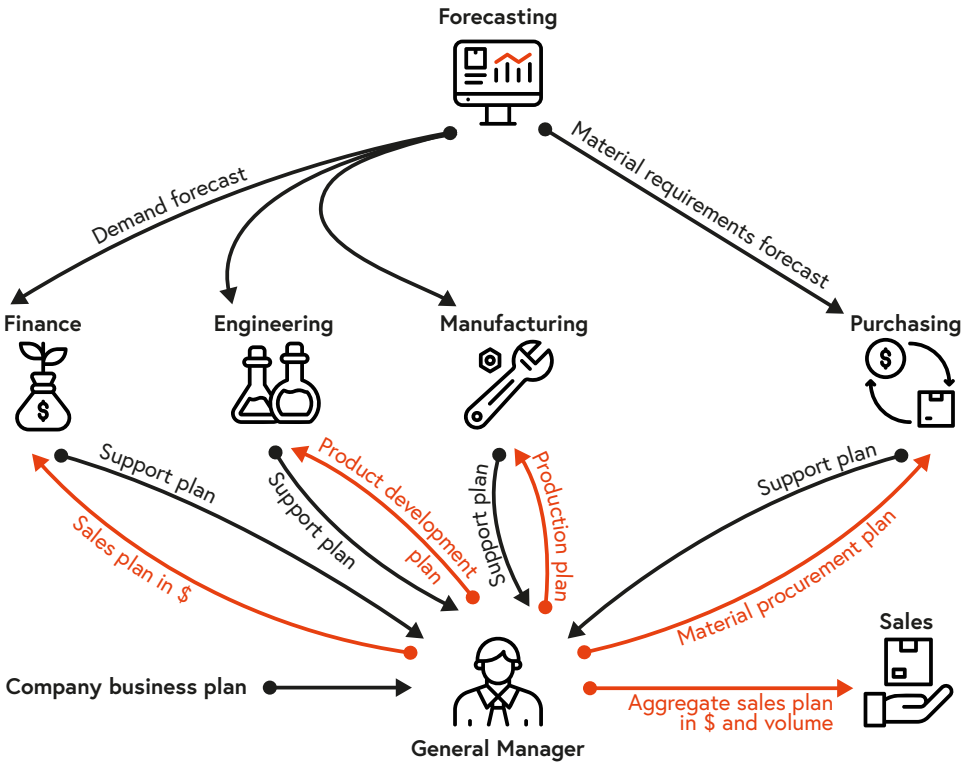


Figure 3.26 Sales & Operations Planning overview

10 Ling, R. C. (1988). *Orchestrating success: Improve control of the business with Sales & Operations Planning.* J. Wiley.

S&OP is started by the issuance of a demand forecast. It is reviewed by all the key managers, who propose to the general manager of the business unit their planned response. The general manager reviews these support plans with the approved business plan and budget, and, after possibly several iterations with the managers, convenes a meeting at which the plan is agreed upon and becomes effective.

In capacity planning, we seek a simple answer to a complex question. If we set capacity too low, we miss sales; if too high, we overinvest. But, except in rare, special circumstances, we never exactly know what the demand will be in the future. There is no general answer that pops out of a mathematical formula. We assume that the past contains information about the future and describe simple analytical tools to retrieve this information and support management decisions. S&OP helps us find better answers. It is a manual process, conducted at the product family level. It relies on each manager's understanding of the capabilities and capacity of his or her organization, and does not require technology beyond an electronic spreadsheet, as shown in Figure 3.27.

Sales & Operations Plan for October, 2020

Product family: XYZ filters
Target line utilization: 95%

Unit of measure: 1,000 pieces
Target inventory: 10 days

	History			Next six months						Following quarters		Year summary		
	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Q2	Q3	12-mth total	Year Rev.	Bus. plan
Sales														
Forecast	50	50	50	50	50	50	50	50	50	150	150	600	$38M	$38M
Actual sales	51	51	49											
Differences														
Monthly	1	1	-1											
Cumulative	1	2	1											
Operations														
Plan	50	50	50	50	50	50	50	50	50	150	150			
Actual	49	51	50											
Differences														
Monthly	-1	1	0											
Cumulative	-1	0	0											
Inventory														
Plan	25	25	25	25	25	25	25	25	25	25	25			
Actual	23	23	24											
Days on hand	10	10	10	11	11	11	11	11	11	11	11			
Line utilization	89%	93%	91%											

Demand Issues
1. Impact of model change in Oct.
2. Upturn expected in Europe
3. ...

Supply issues
1. New machines installed in Nov.
2. Labor issues at supplier X
3. ...

Figure 3.27 Sales & Operations Plan example

Why is the monthly planning process a negotiation between people when, conceptually, it should be a calculation? The scheduling of a cell during a shift is based on hard numbers that are known to production team leaders and supervisors. For example, a machine tool may take one work piece at a time every 53 seconds, a sealant may take 24 hours to cure, and changing these numbers takes a process engineering project.

Logically, the numbers should be usable for monthly planning of an entire product line or even factory, but it would require them to be available in a central master data database, together with the operating policies of each production team and Advanced Planning and Scheduling software (APS). Not only is it technically difficult, but first line managers prefer to retain local control of detailed capacities and only

share with top management a summary in the form of generic units per day or at best a takt time at which they can run a mixed-flow line. As a consequence, the capacity numbers used in monthly planning are soft.

The information in the Master Production Schedule, however, is neither complete nor timely enough to drive production. For example, it can be refined through 10-day orders submitted by dealers a week in advance and subject to limited alteration up to 4 days before a car rolls off the assembly line. The final daily quantities with all option details pass from sales to production control 3 days before roll-off, and the leveled sequences are calculated 2 days before roll-off.

The S&OP meeting serves multiple purposes:

- It helps production managers anticipate, for example, the need for overtime or the opportunity to shut a line down for yearly maintenance.
- It contains information that the Purchasing department can share with suppliers.
- It helps the Accounting department refine quarterly forecasts.
- It informs the Sales and Marketing department about what products should and should not be marketed and sold.

The monthly S&OP cycle often contains the following recurring five steps:

1. **Forecasting**: Gathering information for future demands.
2. **Plan demand**: Actively use forecasts and qualitative information to plan demand.
3. **Plan supply**: Take into consideration current capacity and constraints to plan what can be supplied.
4. **Pre-S&OP meetings**: Make recommendations for the Executive S&OP meeting, often with the use of scenarios.
5. **Executive S&OP meeting**: Discuss and decide on sales and operations.



Figure 3.28 In a monthly S&OP meeting, humans discuss and negotiate numbers

3.6. Takt time

Takt time is a very useful concept for planning capacity in the majority of manufacturing industries. It applies to discrete manufacturing, less so for batch production, not for continuous processes, and not for the manufacture of full custom, one-of-a-kind products. Factories in the process industries have a takt time after the discretization point. For example, beverage companies fill and pack bottles, smelters cast ingots of iron or aluminum, and drug makers make and pack pills. Takt time applies to those processes too.

The exceptions where the takt time concept loses its relevance in discrete manufacturing are where random orders for configured products are directly processed first-in-first-out. It happens in sandwich shops like Subway restaurants, but also in more elaborate manufacturing settings, like custom bag sewing at Timbuk2 Designs or computer assembly at Dell. Instead of being organized around the takt time, Subway and Timbuk2 use the bucket-brigade concept and Dell fixed-station assembly – where an operator assembles all parts in the computer – instead of an assembly line.

3.6.1. What is takt time?

Takt time is one of the most undervalued concepts in manufacturing. Yet it is vital to manufacturing operations. Often confused with process time or with customer demand, the takt time is the time that must elapse between two successive unit completions in order to meet the demand if the factory makes the product one unit at a time at a constant rate during the net available work time. It is a function of both the demand and the available work time. Assembly lines, for example, are designed for a minimum takt time and operated at an actual takt time that is usually longer, and adjusted only in response to major shifts in demand. Takt time is generally considered to be the marching order for a production system.

"Takt" is a common, everyday German word, with the following meanings:

1. A musical bar.
2. A regular beat. For example, trains leave with an 8-minute takt, meaning every eight minutes. It is a natural extension to think of "takt time" as the time between beats.
3. A stroke in an engine. A four-stroke engine is a "Viertaktmotor."

Table 3.2 lists some typical takt times across different industries.

Table 3.2 Typical takt times for some products.

Industry	Approximate takt time
Detergents	1 second
Auto parts	10–30 seconds
Car assembly	1–5 minutes
Truck assembly	5–10 minutes
Commercial aircraft assembly	3-10 days
Offshore oil platform construction	1–2 years

> **Definition**:
> Assuming we complete the product one piece at a time at a constant rate during the net available work time, the takt time is the amount of time that must elapse between two successive piece completions in order to meet the demand.

Hence, takt time is the time available to work on one piece through each operation in a line. Note that takt time differs from *cycle time*, which is the actual time it takes to process the piece. The following equation can be used to calculate the takt time of a product group:

$$Takt\ time = \frac{Net\ available\ production\ time}{Demand}$$

A takt time can be calculated for one product or for a family of products made on the same production line. In all cases, all operations must be designed to fit within the takt time. It is a challenge for a single product, and even more of a challenge for a product family. When doing it for a product family, the line can be envisioned like a ski lift. The time between the chairs is the takt time and it is fixed; the product units are the passengers (see Figure 3.29).



Figure 3.29 The chairlift analogy for takt time

Takt-driven production is an ideal, only approximated in production lines, but all deviations from it introduce waste. It is a global ideal that consists of achieving a local ideal everywhere. Therefore, local shop floor projects that move factories in the direction of synchronizing cycle times with the takt time contribute to improving its global performance.

Box 3.6 **History of the takt time concept**

The "Taktsystem" was developed at the Junkers aircraft company in Germany in the 1920s. The idea was to break down the assembly process into segments of equal duration, during each of which the plane remained in a fixed position. At the end of the segment, all planes were moved to the next one simultaneously. In the US, this is known as a *pulse line*, as opposed to a moving assembly line, with the difference that, in the Taktsystem, the same team of assemblers followed the plane from start to finish. The point of moving the planes from station to station was not division of labor but the logistics of bringing parts, fixtures and tools. Figure 3.30 shows this system in action.

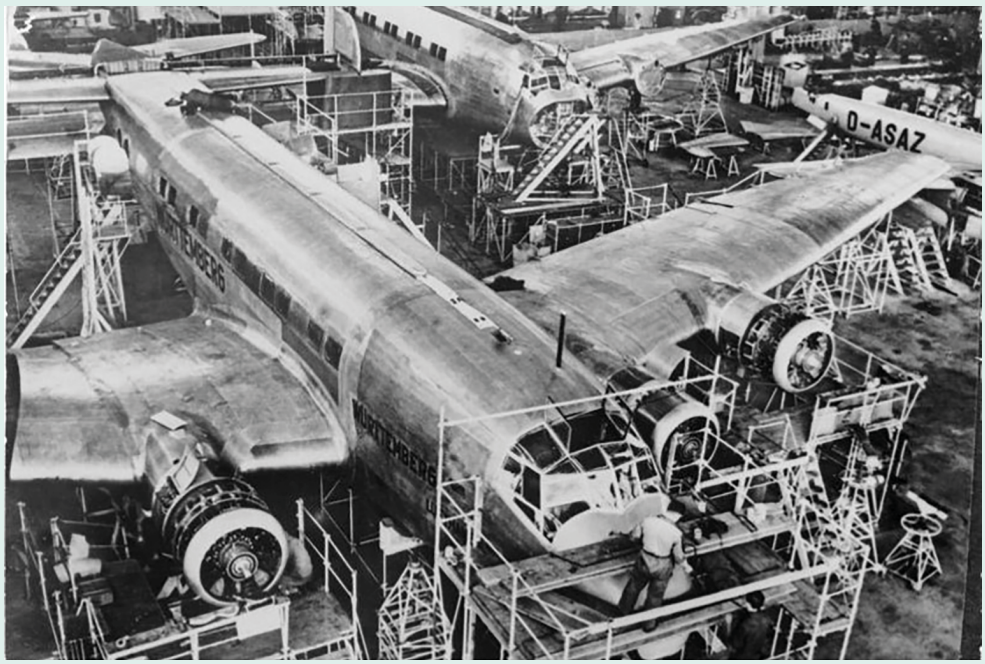


Figure 3.30 A Junkers Ju 90 in the Junkers-Werke Dessau in 1938
(Credit: Deutsche Bundesarchiv Bild 146-1980-003-31)

The Taktsystem reached Japan in 1942, when two Junkers engineers taught it at the Mitsubishi aircraft factory in Nagoya. Mitsubishi implemented it in fuselage assembly. Toyota then learned about it after the war from former Mitsubishi Aircraft engineers. Toyota adopted the Taktsystem, but not the method of having the same team of assemblers follow each unit from start to finish.

By the 1950s, Toyota had adopted and integrated the concept of takt time in its Toyota Production System. As Taiichi Ohno wrote in his book *Toyota Production System* (1988, p. 12): "If a part is needed at the rate of 1000 units per month, we should make 40 parts a day for 25 days. Furthermore, we should spread production evenly throughout the workday. If the workday is 480 minutes, we should average one piece every 12 minutes." By today, takt time is an integral part of the Toyota Production System (TPS) and lean production.

3.6.2. Why takt time matters

An assembly plant manager's response to the question of "What is your takt time?" reveals much about how their production lines are designed and operated. If the manager answers "28 seconds" without hesitation, you can expect one unit of product to leave the line exactly every 28 seconds and every assembly operation to be completed within 28 seconds. On the other hand, some managers will pull out a calculator and say, "Let's see, we make 860 units per shift and work effectively 400 minutes, so output will average one unit every 28 seconds, the rate will vary around this average, and operations inside the line will be run in batch-and-queue mode".

Where the manager knows the takt time, you can expect materials to proceed from one operation to the next in lock step; where the manager does not, like a herd, irregularly, in lumps, and with choke points. The point of using takt time is therefore not to meet demand per se, but to go about meeting it in a particular way. As illustrated in Figure 3.31, producers of washing machines and tractors are usually paying attention to takt time.



Figure 3.31 Assembly of washing machines and tractors have widely different takt times

Stop and Think 3.6
What do you think are the approximate takt times of washing machines and tractors assembly? Why do they have different takt times?

Note that producing to takt time does not directly affect customers. Customers want the products when they need it, not according to clockwork. But, takt-based production affect them indirectly, by improving cost, quality, and delivery performance. To see how it does this, we need to visualize takt-driven production as the ideal situation, as illustrated in Figure 3.32. Each step of the process takes exactly one takt interval to complete, and transfer to the next step is instantaneous. It is easy to see how any deviation from this state of affairs inflates work-in-process (WIP), causes waiting, extra transportation, and other types of waste.

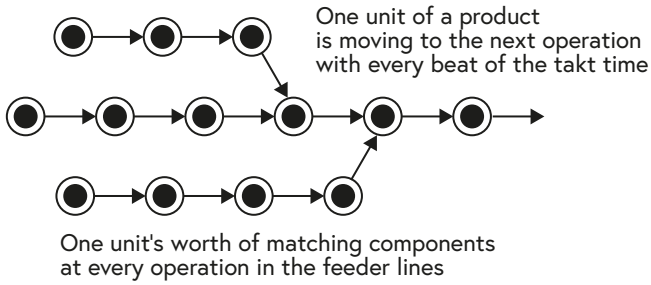


Figure 3.32 The ideal of takt-driven production

In reality, takt-driven production exists at best as an approximation. The technology to perform the entire process is not available in-house, so we have to rely on outside suppliers, who may be thousands of miles away. Some of the in-house resources have to be shared among different processes. In some industries, you even have re-entrant processes, like photolithography in semiconductors, where parts undergo the same process segment multiple times. Finally, the processes are not always in the state of control that our engineers are working to achieve. As a result, we have some testing operations, scrap, and rework. Figure 3.33 shows some of the consequences of these deviations from takt-driven production, all of which create waste.

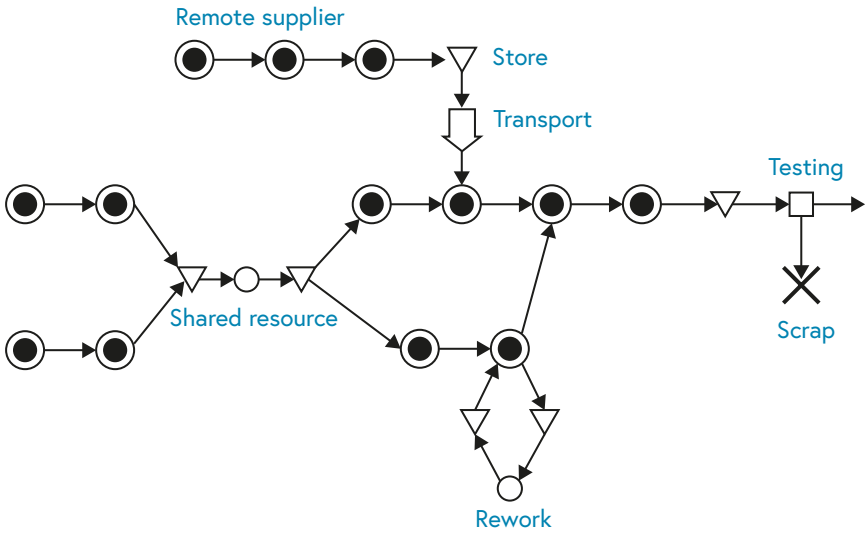


Figure 3.33 Approximation of takt-driven production

That every deviation from takt-driven production adds cost is obvious; that it hurts quality is less so. The reason is that, by lengthening the production lead-time, they also delay the detection of defects, which both increases the number of defective units produced and makes it more difficult to identify the causes of defects. The impact on delivery performance is related to the complexity added to the upstream supply chain. Since it becomes more difficult to ensure the assembly, as the number of needed items grows, so do the chances of shortages in at least some of them. Conversely, the closer we can get to the ideal of takt-driven production, the better we perform in all these areas. From the perspective of the customers' daily operations, takt time is irrelevant. From a longer-term perspective, on the other hand, as your cost, delivery, and quality performance improves, your customers can take share in the benefits, but these accrue over months and years, not days.

Summary

The most essential points in this chapter were as follows:

1. Every manufacturer must do forecasting to anticipate demand using both quantitative and qualitative forecasting techniques.
2. Quantitative forecasting includes time-series analysis and causality models.
3. Past sales patterns should be analyzed with an eye to volumes, variety, and variability.
4. Independent demand of finished goods is forecasted and the dependent demand of parts is calculated from this forecast.
5. Capacity is the rate at which an operation can transform inputs into outputs. In reality, the capacity of a factory is elastic.
6. Sales & Operations Planning (S&OP) is a monthly meeting process where people from across departments seek to align and synchronize sales and production capacity for the following 3 to 6 months.
7. Two extreme capacity strategies are level production versus chase production.
8. Takt time is the amount that must elapse between two successive piece completions in order to meet the demand, assuming we complete the product one piece at a time at a constant rate during the net available work time. The point of using takt time is not to meet demand per se, but to go about meeting it in a particularly efficient way.

Test yourself

1. What is forecasting?
2. What are the five types of variability in a sales pattern?
3. What does the weighted moving average model do?
4. How can you know if you selected the best forecasting model?
5. What are causality models and how can they improve the forecasting process?
6. What is the difference between the capacity strategies "level production" and "chase demand"?
7. What does it mean that the capacity of a factory is elastic?
8. How is Sales & Operations Planning conducted?
9. How does the takt time relate to customer demand?
10. What is the purpose of using takt time in a factory?

Further reading

Baudin, M. (2005). Not-so-basic equipment. *Industrial Engineer*, May issue.

Bartholdi, J. J. & Eisenstein, D. D. (2005). Using bucket brigades to migrate from craft manufacturing to assembly lines. *Manufacturing & Service Operations Management, 7*(2), 121–129.

Brockwell, P. J., Davis, R. A., & Calder, M. V. (2002). *Introduction to time series and forecasting*. Springer.

Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Koutsandreas, D., Spiliotis, E., Petropoulos, F., & Assimakopoulos, V. (2021). On the selection of forecasting accuracy measures. *Journal of the Operational Research Society*. In press.

McCarthy, T. M., Davis, D. F., Golicic, S. L., & Mentzer, J. T. (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices. *Journal of Forecasting, 25*(5), 303–324.

Mentzer, J. T. & Moon, M. A. (2004). *Sales forecasting management: A demand management approach*. Sage Publications.

Overy, R. J. (1994). *War and economy in the Third Reich*. Clarendon Press.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. B., ... & Ziel, F. (2022). Forecasting: Theory and practice. *International Journal of Forecasting, 38*(3), 705–871.

Schoemaker, P. J. (1991). When and how to use scenario planning: A heuristic approach with illustration. *Journal of Forecasting, 10*(6), 549–564.

Wallace, T. F. (2002). *Sales and Operations Planning – The how-to handbook*. T. F. Wallace & Co.

Part 2

Engineering the Factory

DOI:10.4324/9781351110310-5



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

4 Process Design

According to W. Edwards Deming, "If you can't describe what you are doing as a process, you don't know what you're doing." Whether this assertion is applicable to every human endeavor is debatable, but it certainly applies to manufacturing. Processes are the building stones of manufacturing. This chapter describes manufacturing processes in general and outlines principles of process design. It also pays particular attention to the specifics of assembly processes, where direct human labor remains an important resource.

What comes first, the product or the process? Logically, the product comes first. Without knowing what it is you are making, it is impossible to decide how to do it. In reality, however, it is never this simple because the starting point is not a green field. As a going concern, the company has process capabilities, and product designers are expected not to push the envelope too far by designing products outside their area of expertise or that require massive new process development. There are also industries where the common logic is turned on its head, and process development precedes product development. Semiconductor companies, for example, use a test circuit as a "technology vehicle" to develop and test a process. Once the process is sufficiently stable, development teams start designing products to make with it. Large semiconductor companies like TSMC and Intel do both process development and product design.

This chapter describes the different types of manufacturing processes and discuss their key characteristics. It classifies manufacturing processes from the perspective of manufacturing management. The chapter also elaborates on fabrication processes. It leaves out assembly processes, which have been dedicated a separate chapter.

Learning objectives

After working through this chapter, you will be able to do the following:

1. Design a manufacturing process for a new product.
2. Draw a process map for a manufacturing operation.
3. Use different frameworks to classify manufacturing processes.
4. Identify the category of manufacturing process for a product.
5. Use process capability metrics and know their limitations.
6. Calculate the capability of a process and know the limitations of this calculation.
7. Design fabrication processes.
8. Suggest when to use additive manufacturing.

DOI:10.4324/9781351110310-6

4.1. General approach to process design

How should manufacturers move from a product prototype to production? The starting point for process design is the product itself and its volume requirements. Depending on whether the plan is to make just 1, or 1/day, or 1,000/day, the process will be designed differently. The managers must agree on volume requirements, which informs the capacity that must be installed. The volume requirement must be shared upfront with the engineering team in charge of designing the manufacturing processes.

The best source of information about the product is a sample of it together with its drawings and the bills of materials. Usually, the development team makes the product prototype in small quantities that can be made available for the process designers to directly observe its characteristics. It also helps if the product is not "thrown over the wall" from development, some of the process designers have pitched in during development to enhance the product's manufacturability, and some of the engineers from the development team participate in the process design to draw attention to key points. This way, *design for manufacturing* can be improved.

> **Definition**:
> Design for manufacturing is the general engineering practice of designing products in such a way that they are easy to manufacture.

4.1.1. Seven-step approach to process design

Designing a manufacturing process involves a series of increasingly specific decisions leading to investments with consequences reaching far into the future. Starting from the end product, a general approach is as follows:

1. **Set a list of states for the product to go through from start to finish inside the plant.** Working backwards from the end-product is counterintuitive but less error-prone than forward from materials, especially with the end-product in hand. Criteria for choosing where to put these breakpoints include the following opportunities:
   - Testing and partially validating parts and work-in-process.
   - Outsourcing the upstream process of any part.
2. **Identify multiple methods to effect the transitions between states.** The product developers' process is a starting point but alternatives must be considered. These methods are called unit processes. In principle, as long as each unit process is capable, whichever one the engineers select should not affect the performance of the finished product; in reality, their work would be considerably easier if it were always so. Particularly in high-technology industries, there are often interactions between characteristics set in different unit processes that cause a finished unit not to work even though the characteristics observed at each operation are all well within tolerance limits.
3. **Integrate the unit processes into a start-to-finish process that reliably produces working units.** Depending on industry, the challenge of process integration ranges from just stringing together unit processes to reengineering the entire process through failure analysis.
4. **Identify the equipment requirements to execute the process at the pace required.** In some industries, it means selecting machines but it isn't always the case. In machining, for example, the choice of cutting tools and workholding devices – jigs and fixtures – can and should precede the selection of a machine. The core of a machining operation actually is the *control program* (CNC) and a version of it often exists from product development, unless the prototype was made through

additive manufacturing.

5 **Define the roles people will play in the process.** It ranges from executing it with hand tools to monitoring and responding to alarms on an automatic system. This affects the number of people required and their skillsets. As the level of automation rises, the focus for operators shifts from the equipment to the product, as in, for example, applying go/no-go gauge checks between automatic operations. In addition to operators, automatic systems need programmers and the content of maintenance work changes.

6 **Acquire and install the equipment while learning to use it.** This step involves layout decisions. Some new processes require new machines, but not all do. In assembly, for example, the same conveyors and power tools are often used through more than one generation of products. In semiconductor wafer fabrication, on the other hand, a new process usually requires a new plant, with tighter cleanliness specs than the previous one, a new generation of machines and handling systems for larger silicon wafers.

7 **Start and ramp up production.** In almost every case, the migration from making a single unit to making a sample and ramping up daily production uncover many technical, logistical, and managerial problems that must be quickly solved.

This is not a list of steps where each must be finished before the next one starts. There would be a logic to doing it that way, but it would take too long, especially when problems discovered at one stage cause iterations back to an earlier one. Process developers do everything they can to proceed in parallel on all these steps, an approach called *concurrent engineering* or *simultaneous engineering*, fully understanding that some of the developments will not be used. There is a cost to working this way but it pales in comparison with the consequences of a delayed product launch.

Box 4.1 **The Racing Spirit at Honda**

An Ohio engineering company was building assembly fixtures for Honda's plant in Marysville. Honda supplied drawings and instructed the supplier to build the fixture. Before it was complete, Honda provided a revised set of drawings, with features that required the supplier to take apart some of what had been done. Honda agreed to pay for the extra work. This pattern was repeated several times before Honda took delivery of the fixture.

"Why do you do this? It's expensive for you," the supplier asked. The Honda buyer replied: "It's the Racing Spirit."

Early in the company's development, founder Soichiro Honda involved it in motorcycle racing. It was good publicity but racing was also an environment that bred extreme discipline in product development. If the bike is not ready on the day of the race, then all the effort that went into its development is for naught. Doing whatever it takes to be ready on the day of the race is the *Racing Spirit*, and Soichiro Honda wanted it to guide the entire company.

Paying the supplier to redo the fixture multiple times was simply part of "whatever it took" to launch the new model on time. Delaying the launch by one day is giving the competition a 24-hour head start, which Honda engineers considered much worse than paying the supplier for extra labor – they were possessed by the Racing Spirit.

Box 4.2 **Intel's Copy EXACTLY! factory ramp-up system**[1]

Intel develops its next-generation wafer fabrication process in one location. When the yield – ratio of good circuits to the quantity produced – reaches 10%, they turn the development lab into a production plant and start ramping up production. Meanwhile they replicate the facility, the equipment set, the nomenclature and the process programs in multiple locations throughout the world where they want to make and sell products based on this process.

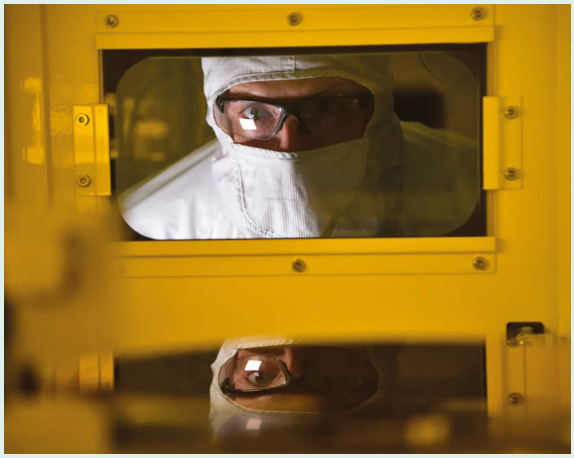


Figure 4.1 Intel copies processes from the lab to the fab *exactly* (Credit: Intel)

The master data and the history of production with all the measurements and test results on products are stored in a shared database and analyzed to identify yield improvement opportunities. Engineering changes are managed globally, so that improvements made in one site are immediately transferred to all the sister sites.

Copy EXACTLY! is Intel's factory ramp-up strategy and has been in place since it was developed for the 0.5-micron wafer generation in the early 1990s. Copying everything helps getting lines up quickly and with the same yield numbers as achieved in the lab.

4.1.2. Importance of human work design

Work design would deserve intensive attention even if operators worked for free. That operator jobs usually have relatively low costs for the firm compared to other jobs does not diminish their importance to the business. The shop-floor operators should be viewed as a competitive weapon. To take an aviation analogy, the cost of airline pilots is small compared to the total cost of owning a jumbo jet. Yet, if you are an airline, how pilots fly is key to your return on this asset.

Where work design is neglected, factories use more people than needed on direct production tasks, and the following happens:

- As in all overstaffed organizations, people fight over the work instead of doing it. Lead times go up and absolute production numbers go down as a result of people being in each other's way.
- Quality suffers as a result of diluted responsibilities. Problems are always somebody else's.
- The morale of underutilized operators plummets, as they feel insecure in their jobs and humiliated that management feels it can afford to waste their time.
- The only way to obtain price concessions is by pressuring suppliers into giving up some of their own profits. On the other hand, if a company can teach its suppliers how to increase productivity, then suppliers can cut prices and increase profits at the same time, and this effect ripples up the

1 McDonald, C. J. (1998). The evolution of Intel's Copy EXACTLY! technology transfer method. *Intel Technology Journal*, Q4, 1–6.

supply chain. Therefore, improvements in production work must precede, not follow, improvements in the supply chain.

These phenomena can be easily observed in developing countries, where factories are often perceived to be a means of "creating jobs for the people." Conversely, the survival of factories in industrialized economies hinges on production work being done by few, well paid, highly qualified people. Companies can afford to pay wages that are ten times higher than their competitors, because the assemblers are more than ten times more productive and their output is of superior quality.

4.2. Drawing process maps

Designing a manufacturing process involves drawing maps at varying scales and levels of abstraction. Process maps express process designs. Maps expose problems and help designers think through solutions. They are also a communication tool to bring new team members up to speed, and input to layout designers. When a process is purely a sequence of operations, or routing, graphics are unnecessary and a simple numbered list is both easier to generate and more effective. Maps are however needed when operations run in parallel, loop, converge, or diverge, which is common in manufacturing.

At the highest level, a process is a black box with incoming materials and outgoing products. As you proceed through the design, you open this box and break into smaller boxes for process segments and then into operations. It is like switching lenses on a microscope for increasing magnification, with each level providing different information and supporting different decisions.

4.2.1. Process maps graphic conventions

Countless graphic conventions have been put forward for process mapping. An early convention was designed by Frank and Lillian Gilbreth in 1921[2] and adopted by the American Society of Mechanical Engineers (ASME) in 1947. Basic process maps in manufacturing often use the symbols shown in Figure 4.2.

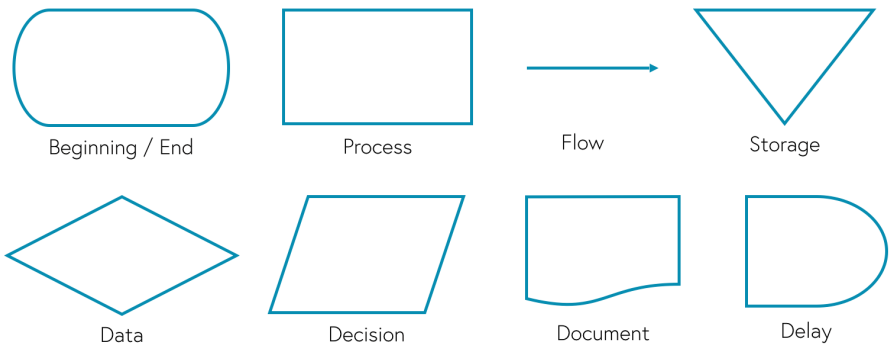


Figure 4.2 Basic process maps symbols used in manufacturing

2 Gilbreth, F. & Gilbreth L. (1921). Process charts. *Annual Meeting of The American Society of Mechanical Engineers*, December 5–9, New York.

The symbols in Figure 4.2 can be found in all common drawing and presentation software, such as for example Google Docs, Microsoft PowerPoint, Word, and Visio. There are also plenty free and subscription-based online process charting tools on the internet.

Sometimes, manufacturing professionals prefer pictograms of their actual products. Figure 4.3 is an example of the process of rolling a slab of lead into sheets that can be rolled up, bringing down the thickness from 170 to 2 mm. Every time you reduce the thickness by half, the slab becomes about twice as long, so you cut it in half and set aside one half before continuing. Once you are done with the first 2-mm sheet, you retrieve the other 4-mm sheet and roll it, then you retrieve the last 8-mm sheet and roll it. This is only one operation but it is the core of the entire shop's process, and it doesn't lend itself to useful mapping with the generic ASME symbols.

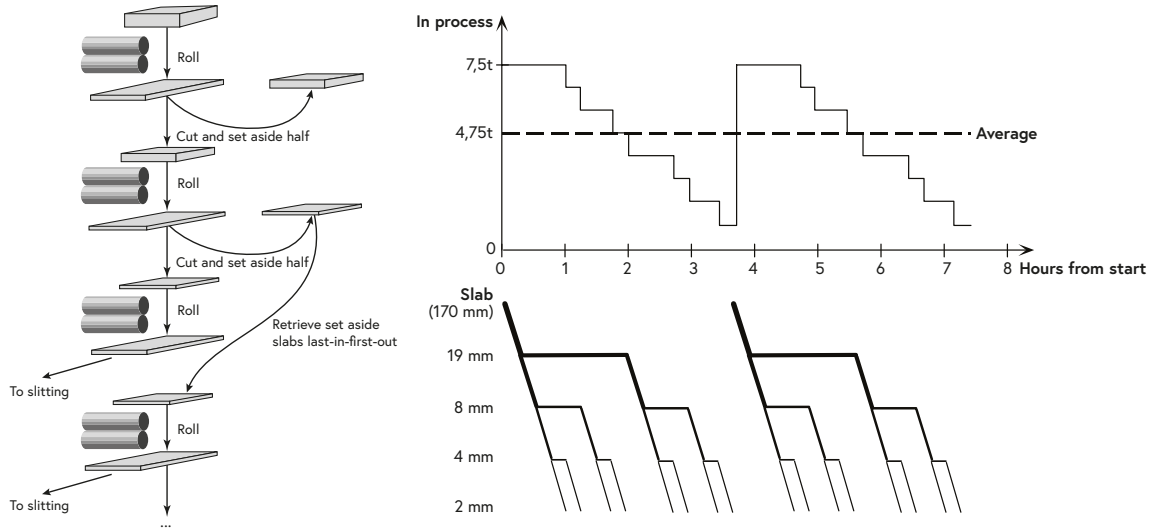


Figure 4.3 Process map with pictograms

In 1961, simulation precursor Jay Forrester at MIT used a more complex but equally abstract convention that was not widely adopted in manufacturing[3] but inspired the field of System Dynamics. In 1989, Shigeo Shingo,[4] in his study of the Toyota Production System from an Industrial Engineering viewpoint, also introduced additional generic symbols that have not caught on either.

Process maps are networks of boxes and arrows where an arrow from A to B should mean that an output of A is an input to B, described as needed by the arrow's label. This is more concise and expressive than what is commonly known as *Business Process Mapping* (BPM), where the arrow just means that A is done before B. There are plenty of software providers for BPM, including integrated modules in the major business software.

The only graphic convention to have caught on is the *Materials and Information Flow Analysis* (MIFA), more popularly known as *Value Stream Mapping* (VSM).[5] Unfortunately, VSM does not support process mapping. One operation box in a MIFA/VSM can contain an elaborate process like "Engine Assembly." The process map shows what happens *within* this box in terms of what machines and people do to

3 Forrester, J. (1961). *Industrial dynamics*. MIT Press.

4 Shingo, S. & Dillon, A. P. (1989). *A study of the Toyota Production System: From an Industrial Engineering viewpoint*. Productivity Press.

5 Rother, M. & Shook, J. (2003). *Learning to see: Value-stream mapping to create value and eliminate Muda*. Lean Enterprise Institute.

workpieces, independently of the facility that is expected to host it or the procedures it follows outside of production. The information part of MIFA and VSM charts is about production control, not process design. It is all about orders, schedules, replenishment signals, etc. By contrast, data flows that you may show in a process map include downloads of process programs to machine controllers and uploads of technical characteristics of the materials or workpieces. In extraction or refining processes, for example, the content of each batch of raw materials varies. As part of the process, the plant analyzes a sample from each batch and tweaks operation parameters as needed.

Other than involving nodes and connectors, process maps don't have a universally accepted graphic language. Checking a map for compliance with a graphic standard is the least relevant way to review it and encourages engineers to generate them as wall decorations rather than useful tools. The maps, however, must be internally consistent, with the meaning of symbols explained in a legend. Unless they are self-explanatory, nodes and connectors should be labeled with clear captions. To avoid ambiguity in busy charts, the *connectors* should also touch the nodes they are linking. A simple process map is effective for a process that fits on a single sheet of paper, but trying to map car making this way would produce an inextricable map covering entire conference room walls. Yet this level of detail is necessary if you want to use the map as a basis for the design of a production line.

Stop and Think 4.1
Can you sketch a quick process map of waking up and leaving the house in the morning using the convention in Figure 4.2?

4.2.2. Overcoming limitations of two-dimensional maps

Maps are two-dimensional. For communications about process designs, it is an improvement over one-dimensional text. A manufacturing operation, however, has more than two dimensions. It has a load size, a process time, a setup matrix, a process capability, etc.

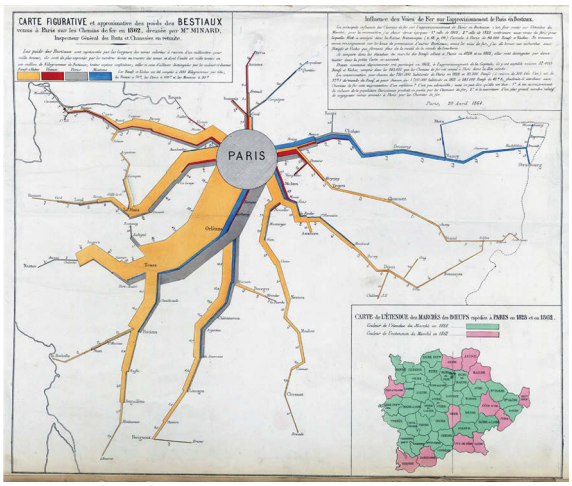


Figure 4.4 Charles Minard introduced linewidth as a measure of volume in flow maps (Credit: Charles Minard / Collections de l'École des Ponts ParisTech)

One technique developed by infographics pioneer Charles Joseph Minard is to make the linewidth of the connections proportional to the volume flowing through, as he did in Figure 4.4 for the weights of beef (yellow), veal (red), pork (gray), and mutton (blue) delivered by rail to Paris in 1864.

The use of different line or arrow width is also usual in manufacturing. In Figure 4.5, it is used in a manufacturing process map in to highlight that the number of defectives sent to Failure Analysis is small compared to the total flow.

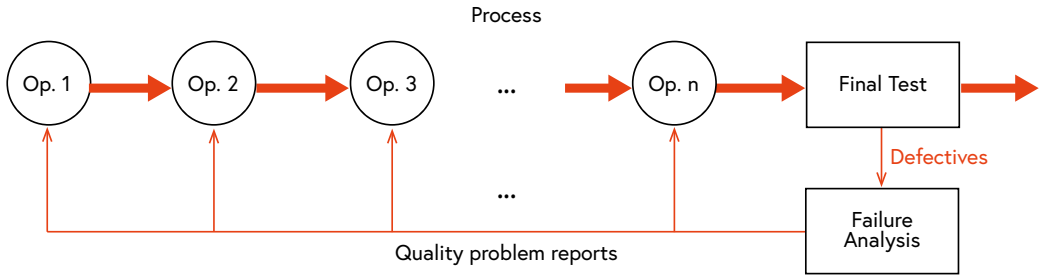


Figure 4.5 Linewidth as a measure of volume in a process map

Given that a failure analysis feeds data rather than materials back to the operations, the map should ideally use a different kind of connector. While, in this simple case, adding 1 more symbol type would not make the chart easier to read, in general, data and materials don't flow the same way and therefore do need different symbols. When a measurement is passed from A to B, it is replicated and A still has it afterwards; when a workpiece moves from A to B, on the other hand, A no longer has it.

Using 3 dimensions rather than 2, the process map could show some operations above others rather than beside them, as chemists do with molecules. On paper, it doesn't have much of an advantage over a flat, top view. It's more difficult to draw and not necessarily easier to read. With a 3D CAD tool on a screen, however, you can turn an object around, flip it, and zoom in and out, and the wide availability of these tools is a potential game changer for process mapping.

4.2.3. Role of simulations

Maps are static. They show what a process is or is intended to be but not how it behaves when executed. Anticipating this in process design involves multiple levels of simulations. With improvements in IT, software simulations for process design have become very popular. Note that there are many different elements that can be simulated, such as material processing, machine movements, work tasks, process designs, and layout. There is also a range of different ways to "simulate," from using data sheets to advanced 3D simulation software:

- **Spreadsheet "simulations."** Checking out what-ifs in an MS Excel spreadsheet by adjusting parameters and refreshing a workbook is often called "simulation." All it simulates, however, is the behavior of a mathematical function that may have a tenuous connection to a physical reality.
- **Simulation software for the physics or chemistry of the process.** These are used upfront by process engineers to set process parameters and, in advanced processes like additive manufacturing, as part of the control system during execution (upper right photo in Figure 4.6).
- **Discrete-event simulations** use software to simulate the events that occur in production, like release work to the process, starting and completing operations, transferring work between operations, and experiencing random fluctuations in demand, quality, or in machine performance. The simulated activity is summarized in performance metrics. With the development of IT, these models have gone from 2D flow diagrams interfaces to full-blown digital copies of factory processes. Although it is not *logically* required, discrete-event simulation software products usually show animations of the process execution and have a graphic interface to develop models with the same symbols used in animation. Without animations, the development of a simulation model requires coding and the visualization of the model is a challenge. At the time of writing this book, there are plenty of providers, including Arena from Rockwell Automation, Plant Simulation from Siemens, FlexSim, AnyLogic, Simul8, Visual Components, and many more.

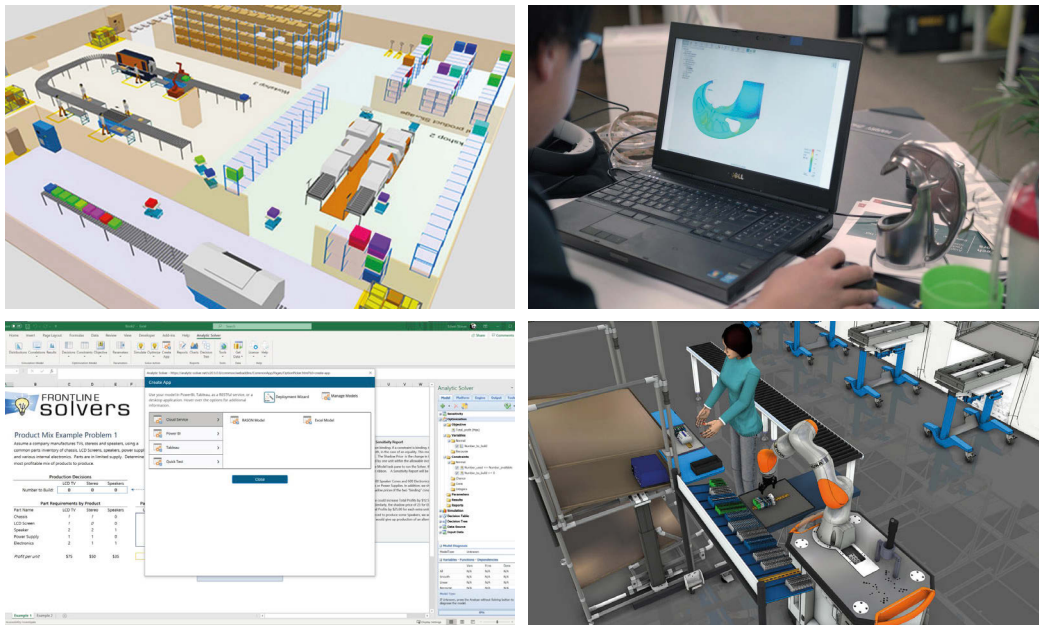


Figure 4.6 Many different forms of software simulation for process design (Credits: AnyLogic, Autodesk, Microsoft, and Visual Component)

Simulations do not have to be done in computer software. When the point is to include the operators in designing processes, physical simulations can be more engaging and spark creativity:

- **Board game simulations.** In a board game simulation, scale models of work pieces, machines, and other objects used in production are used. The board game is used to play production and logistics processes, and collect statistics on lead time, productivity, and quality. It is labor-intensive to set up and the amount of data generated is limited. This method is widely used in training but is also useful in process design, as playing the game reveals design flaws and raises questions in the minds of players.
- **Cardboard engineering.** Unlike the board games, cardboard mockups are full-size renditions of machines and workstations, and their purpose is to allow operators to anticipate the experience of the process and provide feedback on its design. This method has been used in manufacturing for decades, and been adopted by hospital architects for exam and operating rooms.

4.3. Classifying manufacturing processes

There is not one universal way to classify manufacturing processes. In this book, we separate different process types using the following useful typologies:

- **Convergent versus divergent processes**: The form of the process
- **Discrete versus process manufacturing**: The output of the process
- **The product-process matrix**: Volume and mix characteristics of the process
- **Process types**: What the process does to the input material

4.3.1. Convergent versus divergent processes

Assembly represents the typical convergent manufacturing process, but any process that attaches two or more different parts are convergent. Examples include mixing of ingredients in pharmaceutical industries, welding of steel structures, and assembly of bikes. Other processes take a piece of material and splits it into several outputs. Examples are food processing like cutting meat or churning milk, or petrochemical processes like refining oil and gas into several output products.



Figure 4.7 Convergent bicycle assembly (left) and divergent meat cutting processes (right)

Since not all manufacturing processes increase or reduce the number of parts that goes into the process, a more nuanced classification is also useful. Some processes take a piece metal and bend it, carve it, heat treat it, grind it and paint it into a product. Tools and add-on materials like paint are consumed along the way but, essentially, the process turns 1 input into 1 output and is defined by a sequence of operations, or *routing*. It can legitimately be called a *one-to-one* process. An assembly line dedicated to 1 product, by contrast, is a *many-to-one* process, and processes that split 1 input into many different outputs like refinery or recycling are *one-to-many* processes. Finally, mixed-flow assembly is a *many-to-many* process. An actual manufacturing process may break down into multiple subprocesses that each match one of these patterns. Table 4.1 lists examples and key challenges associated with each pattern.

Table 4.1 Convergent and divergent processes

Category	Examples	Key challenges
One-to-one	Machining Fabrication Conservation	Process capability Material quality
Many-to-one	Assembly Mixing	Supplier management Logistics Assembly methods
One-to-many	Refining or recycling Slaughterhouse operations Smelting	Product mix determined by incoming materials
Many-to-many	Mixed-flow assembly Configured products	Mistake-proofing Quality assurance

4.3.2. Discrete versus process manufacturing

At a high level, it is common to separate discrete versus process manufacturing. It's an established terminology but unfortunate because the word *process* applies both to all of manufacturing and to a subset of it. It's like using *New York* for an entire state and a city within it. The context usually resolves the ambiguity but confusion remains possible.

Many manufacturing processes, in the broad sense, can easily be defined as either discrete or process manufacturing. For example, Figure 4.8 shows an excavator produced by mostly discrete manufacturing and a chemical liquid produced mainly by process manufacturing. Discrete manufacturing processes produce discrete units. They cannot be split into two halves of the same product. Process manufacturing, on the other hand, produce indistinct units. You can split a kilogram of sugar into two halves of sugar, or a pint of beer into two half-pints.



Figure 4.8 Process manufacturing of a chemical liquid and discrete manufacturing of an excavator

Finished products like excavators, bikes, cars, phones, furniture, white goods, toys, apparel, aircraft, and so on are results of *discrete manufacturing processes*. Discrete manufacturing processes have the following characteristics:

- Products are assembled together based on a tree of subassemblies and purchased materials called bills of materials (BOM).
- Products and parts are counted in units.
- Each unit can be assigned a serial number.
- Each process step is an identified, traceable operation.

Chocolate bars, sodas, medicines, gasoline, paint, paper, steel, and many more are considered *process industries*. Process plants mix, separate, ferment, or otherwise transform materials through chemical reactions. The processes are described through formulas and ingredients rather than bills of materials and single units. The manufacturing processes for these different products have the following characteristics in common, all of which is illustrated by the cheese manufacturing in Figure 4.9:

- Production of an intermediate product in bulk.
- A discretization point where the bulk product goes into containers. For example, textiles are cut into small pieces; molten chocolate, cast into candy bars; toothpaste, filled into tubes. Where the bulk materials are discretized depends on the industry. If this happens at early stages, the industry can be termed *semi-process industry*.
- Bulk quantity measured in volume or weight.

- Production in batches. Traceability is often necessary, from a finished container sold to consumers back to the batch of bulk that was dispensed into it. A tube of defective toothpaste comes from a batch of thousands of liters and all the tubes from this batch must be recalled.
- Supply of raw and semi-finished material as liquids or slurries through pipes and of powders or pellets through belts or by gravity, sometimes aided by vibrations or pressure.
- Most production operations in process industries are dependent on time. In discrete manufacturing, an operation can often be started at one time, interrupted, and continued at another time. Many chemicals deteriorate over time and food rots. As a consequence, these materials need to be processed promptly and processes are not easily interrupted.
- The factories are asset-heavy and automated. Human intervention is often limited to feeding raw and packaging materials, alarm monitoring, changeovers, and maintenance. We tend to associate process industries with massive facilities, like oil refineries. In industries like cosmetics or pharmaceuticals, however, the process part of the plant is often small compared to the packaging areas and the warehouses. Regardless of its size, however, it is the core of the manufacturing process and least likely to be outsourced.
- Many process plants have long changeover times, particularly when they involve cleaning pipes and kettles between products.
- The transformations occurring during the process are often invisible for the workers. In extreme cases, operators do not even see the material throughout the whole production process, for example, when petroleum is being refined in petrochemical plants, raw materials arrive in pipes and leave in pipes or tanks.

There are, however, limitations to the dichotomy between discrete and non-discrete manufacturing that we will come back to in this chapter. For example, it is not clear how the main production processes of semiconductors, biomedical products, or 3D-printed parts fall into any of these two categories.



Figure 4.9 Manufacturing of Swiss cheese in Maison du Gruyère (Credit: Roman Babakin / 123RF)

Stop and Think 4.2
Make assumption and classify – as mostly discrete manufacturing or mostly process manufacturing – a typical factory manufacturing the following ten products: Jam, guitars, cheese, steel ingots, sofas, computers, paper, salami, medical eye drops, pizza.

4.4. The Hayes-Wheelwright product-process matrix

Separating process types and product types, Robert H. Hayes and Steven C. Wheelwright suggested the classification of manufacturing processes shown in Figure 4.10.[6]

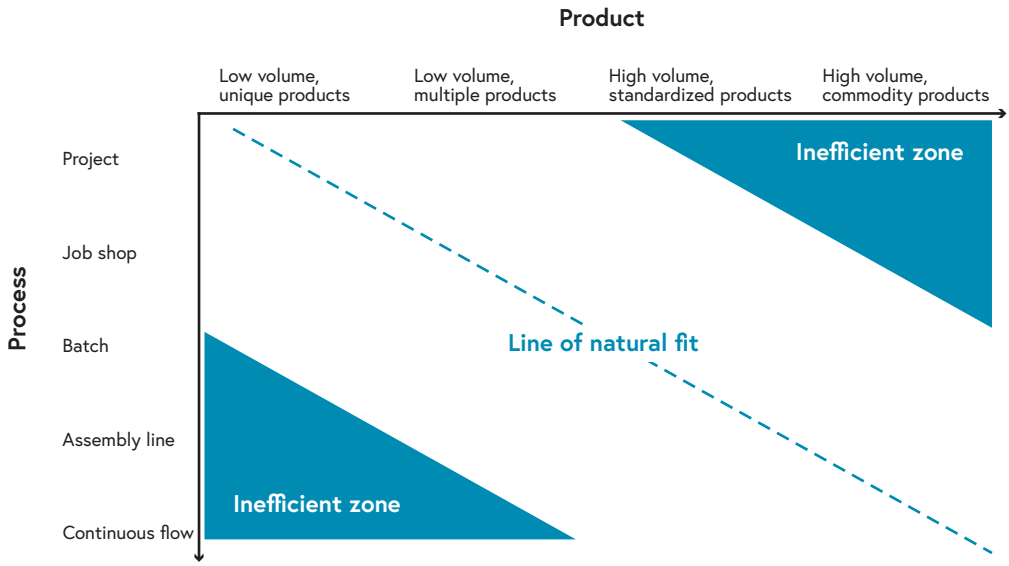


Figure 4.10 Hayes and Wheelwright's product-process matrix

The essence of the Hayes and Wheelwright matrix is that the process must fit the product characteristics. Another key point in this model is that efficiency can be increased by moving right-downwards in the product-process matrix, but it should of course not be done on the expense of effectiveness. It also suggests that there is a natural fit between certain process designs and product characteristics, as illustrated with the diagonal "line of natural fit" in Figure 4.10. Further more, the matrix suggests that moving processes towards down right will increase productivity.

This matrix is popular in the business literature, often reproduced, and widely taught. While it is helpful for gaining a quick overview of more efficient product-process designs, it is, however, simplistic. It has, for example, the following limitations:

- It ignores demand variability. A low-volume, unique product may be perfectly suited for line production if the volume is stable and you have equipment that is appropriate for it, meaning old, slow, and inflexible but with the requisite capability and just enough capacity. Many machine

6 Hayes, R. H. & Wheelwright, S. C. (1979). Link manufacturing process and product life cycle. *Harvard Business Review*, *57*(1), 133–140.

shops have such machines on standby, accumulated over decades, that can be pressed into service with new jigs and fixtures, an overhaul of their moving parts and a control system retrofit, at a fraction of the cost of new machines.

- It restricts the line concept to assembly. Not all line processes are assembly. In machining, you have transfer lines and cells with one-piece flow through a sequence of steps. Ditto in fabrication processes like electroplating, painting, PC board fabrication, solar panel fabrication, semiconductor wafer fabrication, etc.
- The product categories include "standardized product" and "commodity products" but not *configurable or customizable products*, among which are the largest and most expensive manufactured products, from computers to cars and airplanes.
- It states that a project process is the best fit for products that are only produced one time or very few times. As explained in Chapter 1, projects do not really qualify as manufacturing and they do not necessarily take place in factories. For example battleships, oil platforms, and cruise ships are constructed in dockyards. Bridges and buildings are mainly constructed on site. As indicated with the word "project," the processes are often not repeatable and cease to exist when the project is completed. Project management is a rich discipline on its own.
- A job-shop, as opposed to flow-shop, is characterized by the absence of fixed routes. The machines are laid out in farms by type. In a machine shop, all the lathes are together, and so are all the milling machines, all the grinders, etc. A workpiece may visit all these farms in its own sequence, and be processed on any of the machines within each farm. The physical location of a workpiece in a job-shop tells you nothing about where it is in its process, which is why it has a "traveler" attached. Traditional machine shops are organized as job-shops even in high-volume/low-mix, stable production both in automotive and aerospace and the white zone for the job-shop concept in the Hayes-Wheelwright matrix extends all the way to "High-Volume Standardized Products."
- The placement of the assembly line as a category between batch processing and continuous flow is problematic, for the following reasons:
  - Not all assembly work is performed in assembly lines. Fixed-station assembly is still practiced, for example, for large products like port cranes, small products like locks and watches, or for one-of-a-kind, custom products.
  - Not all work done in one-piece flow lines is assembly. Transfer lines and machining cells, for example, are lines in which each operation removes material from each workpiece.
  - A continuously moving assembly line qualifies as continuous flow.
  - The assembly line is most efficient for stable volumes of standard products but also works for products that are configured or customized to some extent.

The Hayes-Wheelwright product-process matrix is also challenged by developments like additive manufacturing, the best known special case being "3D printing" for plastic parts. Other forms of additive manufacturing include melting tiny steel balls in a tub along a path with a laser beam. Additive manufacturing promises to enable manufacturing of high-volume, multiple products conducted most efficiently in a single operation.

A complete, generic set of rules to choose one type of process design based on demand volume and mix cannot fit on a simple matrix or chart. It may be more effectively presented as a decision tree with many branches, leading to leaves that are recommended structures for production processes. In a specific manufacturing business, on the other hand, there are fewer options and this kind of matrix or chart can be a useful tool to support strategic decisions, both in developing options and in communicating the final choices.

Box 4.3 **High-volume production of roller bearings in a job-shop**

It is not clear that a generic solution to the classification of processes can be represented in the form of a matrix with a handful of rows and columns. In particular, these matrices ignore variability, the fluctuations, trend, seasonal, and cyclical variations in demand that must be a consideration when designing a process. A single company's analysis of the appropriate response to the specific demand mix of a business unit may be more enlightening, as directly connected to design decisions and directions in which to pursue improvement.



Figure 4.11 Bearing plant in Sandusky, OH, built after World War II (Credit: Sandusky Library)

In the mid 1980s, the New Departure Hyatt plant in Sandusky, OH, shown in Figure 4.11, employed several thousand workers making 8 varieties of roller bearings for GM in a job-shop. Rows of identical machines extended beyond what the eye could see, as the end of the line was lost in the oil mist. After completing an operation, the workpieces went onto overhead conveyors and were stored in spiralling conveyors while awaiting transfer to the next line of identical machines. The concept is shown in Figure 4.12.

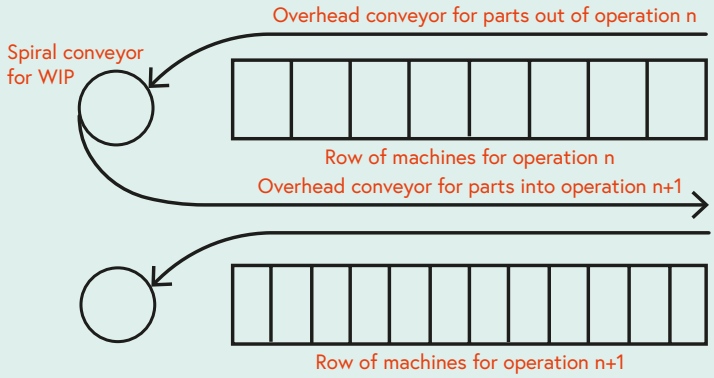


Figure 4.12 The Sandusky job-shop concept

Compared to flow lines, this seems a complex system. Its network of overhead and spiral conveyors, in particular, is an additional point of failure. During operations, the rattle of parts coming down the spiral conveyors generated more noise than the production machines.

After a chain of ownership changes, the Sandusky plant shut down in 2016. In the 1980s, another bearing manufacturer, SKF, launched the "channel concept" for bearing production, which they describe as based on the same principles as line production. They are still using it in 2021.

4.4.1. A comment on job-shops versus line design

Assume a two-step operation process with a volume requiring 3 machines of type X for Operation 1 and 10 machines of type Y for Operation 2. In a job-shop, you can meet the demand with exactly these numbers of machines and keep all 100% utilized. To link the two operations in flow lines so that the output of Operation 1 feeds directly into Operation 2, you need 7 more machines of type X, and they will be used only 30% of the time. And the difficulties are compounded when the machines of type X and Y have different load sizes and maintenance requirements.

Besides the value of flow, there are other arguments why moving towards line production may still be a good idea. For some machines, like a drill press feeding a machining center, availability trumps utilization. It doesn't matter if it is used 5% of the time; what does matter is that it is always ready to feed workpieces to the machining center that you want to keep busy. This argument however only applies when converting from a job-shop to a flow line with the same equipment. If you set out to design a flow line, you look for machines that process one piece at a time, in times that are as closely matched as possible. If such machines are not commercially available, you retrofit existing machines or build your own.

An argument in favor of job-shops is the convenience of facilities management. Machines require power, consume fluids delivered through pipes, waste materials and scrap that must be taken away, and fixtures or dies that must be stored nearby. Operators often require special training to operate a type of machine safely and effectively. Some machines require operators to wear personal protection like ear plugs, hard hats, safety glasses, shoes, or gloves; others require operators to wear special clothing to protect the products from contamination. This is easier to do with collocated identical machines than with each individual machine integrated in a flow line. It is often a legitimate concern. You don't see a diecasting machine integrated with an assembly line, because, among other things, it would require you to move molten aluminum on the assembly floor. On the other hand, as in the case of roller bearings (Box 4.3), a materials handling system adds complexity.

Technology changes sometimes remove the obstacles that hinders implementing a line flow. Heat treatment by induction hardening, for example, can be done one small piece at a time and integrated in a machining line. Figure 4.13 shows how the old oven technology (left) required a special facility, while the modern induction hardening technology (right) works one piece at a time and can be integrated in a flow line.

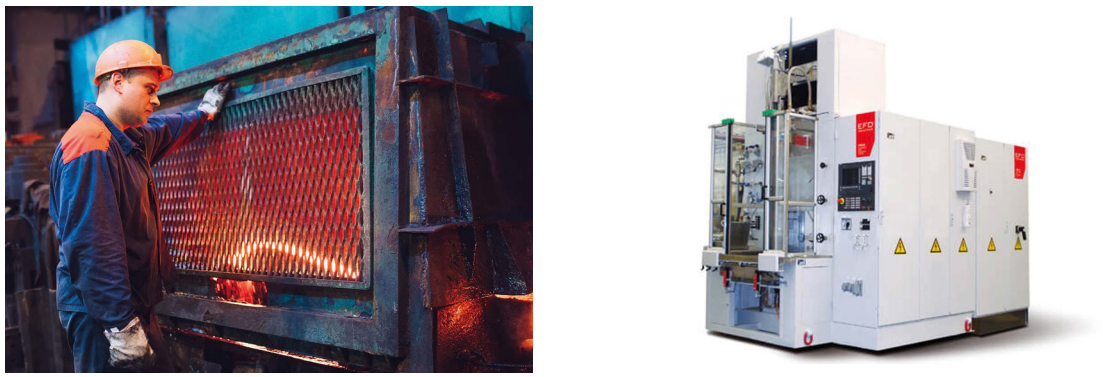


Figure 4.13 Old heat treatment technology (left) (Credit: David Tadevosian / 123RF) versus new induction hardening machine (right) (Credit: EFD Induction)

Figure 4.14 illustrates old MIG ("metal inert gas") or TIG ("tungsten inert gas") welding versus modern friction-stir welding. In factories that do both welding and other kinds of assembly work, the welding shop is traditionally separated because it produces sparks that require eye protection and smoke that requires ventilation. Robotic friction-stir welding produce neither sparks nor smoke and can be more integrated with the rest of the assembly process, as can be seen at Tesla in Fremont, CA.



Figure 4.14 Old MIG welding (left) and modern (right) friction welding technology (Credit: KUKA Group)

Most machinists in large job-shops have a single skill. They can run lathes, or milling machines, or grinders but not more than one type. In flow-shops, on the other hand, operators run multiple machines of different types, requiring them to be multiskilled. In advanced factories, the job-shop concept is still relevant for the production of "strangers" or low volume and high variety products, and in tooling shops that make fixtures and jigs for internal use. Repair shops are often job-shops, as are custom woodworking shops making, for example, display structures for museum exhibits. These job-shops are small, with no more than a handful of workers, who are multiskilled.

Outside of production, R&D labs are job-shops where each job is an experiment and the equipment is operated by the experimenters themselves, not by permanent operators attached to the lab. The experimenters get help from lab technicians for the use of machines or instruments they are unfamiliar with.

4.4.2. A comment on "batch processes"

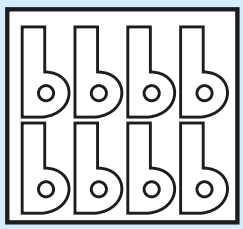
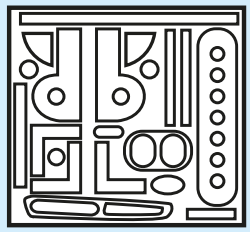
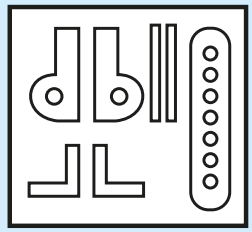
Batch is an often used but also ambiguous term. There are several different concepts that need to be distinguished but are all called "batches":

- The load of a machine is the set of parts it processes at the same time. They may or may not be identical; a stamping press can cut 20 identical pieces from a sheet of metal, or cut different pieces.
- A production run is the number of units made between changeovers.
- A transfer batch is a set of parts moved between two operations. A production run may be 20 bins but each bin can move to the next operation without waiting for the others.

An operation run on a machine with a load of more than one identical workpieces is commonly called a *batch process*. There is no standard term for the simultaneous processing of multiple parts for *different* items, which is pursued for two reasons:

1. To increase capacity. When parts are loaded into a cavity, as in an oven, for example, you can fit more in a single load by mixing large and small parts.
2. To perform "one-set-flow" by processing together a matching set of parts eventually going into the same product. This is illustrated in Table 4.2.

Table 4.2 Different types of machine loads

	Batch of identical units	Nesting to fill slab	Kit of matching parts
Concept	Cutting a regular grid of identical parts from one blank	Filling a blank with as many parts as possible	Cutting a set of parts that go into the same assembly from one blank
Illustration			
Motivation	Comply with equipment constraints	Minimize scrap and maximize machine utilization	Match consumption at assembly
Example	Most sheet metal plants	Patterns for sewing or composite material mash	Stamping multiple parts for cars at once

Growing crops in fields is a batch process that humans have run for 10,000 years. As a consequence, batch processing is the default mode of operation that individuals and organizations adopt when designing processes. The idea of flowing work through multiple operations, on the other hand, dates back to the Industrial Revolution, more than 250 years ago. As an approach, it is surprisingly still perceived as disruptive.

Mixing different items in machine loads is a more complex operation than batch processing, but the extra effort is minimal when compared to the value of extra capacity or of a smoother flow. The control software of batch processing equipment does not always support it.



Figure 4.15 Batch processes are usually inevitable in heavy process industries (Credit: Zhao Jiankang / 123RF)

Sometimes technology is dictating the use of batch processing. For example, steel bar manufacturers use large and heavy tools to cast several bars at once from a continuous batch of liquid steel (Figure 4.15). Other times the processing characteristics, such as long cycle times, make batch processes the only way to make many units in a short time.

Box 4.4 **Process design strategies at Omron**

In 2002, Japanese electronics company Omron published the classification of its own sensing products shown in Figure 4.16. The x- and y-axes are numeric and the curve is drawn with the unstated assumption that you are dealing with constant aggregate volume. Note that the ranges of applicability of the different approaches overlap.

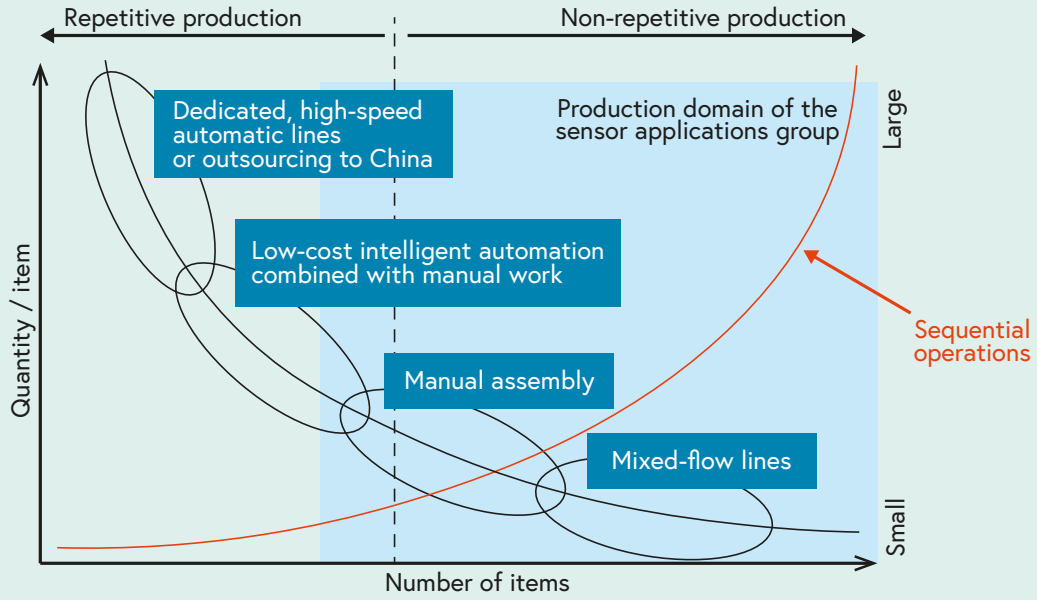


Figure 4.16 The Omron volume–variety chart

Omron's key observation was that, as variety increases and quantity/item decreased, the amount of "sequential operations" skyrocketed, meaning that each operator had to perform increasingly long sequences of operations, and each one so rarely that operators cannot memorize them. The result was that operators were challenged to (1) keep track of the work that had already been done, and (2) find out what needed to be done next. In a breakdown of the operators' use of time in these sequential operations, hesitation accounted for 50% and was the top category.

The conclusion Omron drew was that they needed a prompting system for one-of-a-kind products in order to eliminate operator hesitation and errors. It became what they called their "digital yatai," a concept that has evolved since from triggering the display of information on screen to projection mapping onto the workpieces themselves.

4.4.3. Manufacturing process types

Because manufacturing is about changing materials into products, it also makes sense to classify processes based on what they do to the materials. One can separate ten process types as follows with examples:

1. EXTRACTING
   - Refining
   - Smelting
   - Filtration
   - Distillation
   - Reverse osmosis
2. SHAPING
   - Molding:casting: mechanical formation of part from powder or liquid
   - Forming: mechanical deformation of solid part as in extrusion, roll forming, press braking, stamping, forging, crimping
3. CHANGING MATERIAL PROPERTIES
   - Heat treatment of metals
   - Firing of ceramics
   - Surface oxidation
   - Biological treatment: Fermentation, cell-growth
   - Cooking: Blanching, boiling, roasting, braising, frying
   - Preserving: Freezing, freeze-drying, drying, salting, smoking, pickling
4. REMOVING MATERIALS
   - Machining
   - Etching
   - Sandblasting
   - Shot peening
5. ADDING MATERIALS
   - Coating
   - Chemical vapor deposition
   - Sputtering
   - Additive processes
6. DISASSEMBLY
   - Detaching
   - Slaughtering
7. ASSEMBLY
   - Joining
   - Fastening
   - Labeling
8. SORTING/FILTERING
   - Removing defectives from product flow
   - Binning integrated circuits by speed, or eggs by size
9. CHECKING
   - Inspection/quality control
10. PACKAGING
    - Wrapping
    - Boxing/bagging/bottling
    - Labeling

At the beginning, raw materials may be dug or pumped out of the ground, like coal, ores, rare earths or crude oil, which represent *extraction* processes. They may also be byproducts of human activity, like scrap metal, electronic waste, or recycled plastics. Or they may be harvested from land or sea. They almost always require some form of preparation to be usable in the manufacture of products, including sorting, washing, refining, smelting, etc. One key characteristic that is common to these processes is that the composition of the raw materials and even its availability varies in ways the manufacturer cannot control and, as a consequence, so does the mix of outputs. The metal content of ores or electronic waste will vary with each delivery, and so will the quantities of metals extracted from it.

At the opposite end of manufacturing activity, nearly all products must be packaged in some way before shipment. While the necessity of *packaging* is recognized, it is usually an afterthought in product design. One exception is IKEA, where the ability to fit parts into flat packs is a constraint on product design.

Between raw materials extraction and finished goods packaging, many processes are networks of operations that draw on multiple technologies, requiring equipment that varies in cost, size, skills required, and environmental impact. Many operations can be collocated but some cannot be. Lathes, milling machines, drill presses, and other machine tools can be side-by-side, as can assembly stations along a line. On the other hand, die-casting and assembly operations do not belong together. It is therefore possible to categorize the eight remaining process types into:

- Fabrication processes: combing processes of *shaping, changing material properties, removing materials, adding materials*
- Assembly processes: *disassembly, assembly*
- Selection processes: *sorting/filtering, checking*

"Fabrication" means transforming materials through a *sequence of operations*. The key descriptor of an assembly process is its bill of materials, for a fabrication process, it is its routing. Even when focusing on sequences of operations, the issue of materials is never completely absent. Machining processes remove materials from workpieces, which must then be disposed of and recycled. Other fabrication processes involve the diffusion of some materials into the workpiece, the growth of a thin film on its surface by means of a reaction, or the deposition of a coating. Conversely, assembly processes have routings too. Fabrication and assembly processes often coexist in the same plant. The key point is that the focus of process design in fabrication is on the transformation itself and not on the availability of the materials.

On the ten specific manufacturing processes, books are available that provide all the necessary background and detail, which is outside the scope of this textbook. There are plenty of textbooks covering manufacturing processes. On extrusion, for example, Chris Rauwendaal[7] has authored *Understanding Extrusion*, for technicians or managers, and the much deeper *Polymer Extrusion* for engineers. A trade magazine like *Extrusion International* can supplement this information with industry trends, and a journal like *Materials Science and Engineering* with the latest research.

7 Rauwendaal, C. (2019). *Understanding Extrusion*. Hanser/Gardner Publications. / Rauwendaal, C. (2014). *Polymer Extrusion*. Hanser/Gardner Publications.

4.5. **Process capability**

In principle, a manufacturer wants fabrication processes to be fully capable, meaning that they consistently produces output that works as expected. This *true characteristic* frequently cannot be observed directly, because it would require destructive testing – as in eating a cake to verify its taste – or would take too long – as in checking that paint in a can lasts for 10 years. The practical assessment of process capability is therefore based on *substitute characteristics* in the form of observable, true/false attributes and numerical parameters with tolerances. With modern data science, substitute characteristics can be selected by applying for example logistic regression or neural networks to historical data and estimate the probability that a unit is good from the values of substitute characteristics.

4.5.1. **Tolerances**

Tolerances are set to guarantee that units with any substitute characteristic outside the tolerance intervals won't work. There generally is no way with substitute characteristics to *assure* that a unit actually works. They only make it more likely that it will. Using these substitute characteristics, however, is all a manufacturer can do on a routine basis. It has to be supplemented with observations of the products' performance in the field but this is beyond the issues we want to cover in this chapter.

> **Definition**:
> A tolerance is defined by the range of variation permitted in maintaining a specified dimension.

Process capability is usually not an issue in machining because current machine tools are well able to hold the tightest dimensional tolerances required. Defects are not caused by lack of accuracy but by discrete events like tools breaking or human errors like using the wrong program. In other fabrication processes, competitive pressure often forces manufacturers to start production before engineers have a chance to make the process fully capable. Semiconductor fabrication processes, for example, start volume production with yields as low as 10%; automotive processes like the diecasting of transmission cases, with only 70% of non-leakers. Then, the improvement of process capability is called "*yield enhancement*" and its pursuit is a race among competitors. This is a chronic condition is high-clockspeed industries like semiconductors where, if your processes are mature, your products are obsolete.

In operations, this implies an emphasis on cycles of learning, each involving data acquisition on parts in-process and at the end of the process, data analysis, and engineering changes. The shorter the lead time from start to finish, the more cycles of learning the organization goes through for the same volume, and the faster yields go up.

If you interpret the yield as the probability that a completed unit is good, then the output quantity is random. If the yield is 80% and the quantity ordered is 800, starting 1,000 units will result in 800 good finished units to ship at the end, *on the average*. The process will overproduce or underproduce most of the time. For a *Runner or a Repeater*, these fluctuations can be covered by a finished goods buffer; for a *Stranger*, by overbuilding enough to ensure that at least 800 units will be available, with the near certainty of having more, in an inventory of slow-moving finished goods.

4.5.2. Measuring process capability

When is a process capable? The literature on quality defines process capability as a metric that compares the variability of its output with tolerances. There are, in fact, two different concepts:

- The ability of a process under nominal conditions to consistently produce output that meets expectations.
- The means of assessing it.

A process running under normal conditions has the materials and tools it needs and skilled operators. There are no broken tools, leaks, blown fuses or cracked columns, or unqualified operators filling in for missing colleagues. Yet, under these perfect conditions, the process still doesn't put out exactly identical units every time, and the process capability is a rating of this variability against specified expectations.

Some of the discussions in this section reference the *gaussian distribution*, also known as the "normal" distribution, *standard deviations*, and *analysis of variance* (ANOVA). The reader is expected to be familiar with these concepts.

Cp and Cpk

The nature of the expectations and the means of assessing whether the process meets them consistently vary with each case, and the usual metrics found in the literature, C_p and C_{pk}, are a simple, special case, well defined only for single, substitute characteristics that are measured variables. If you search the internet for "process capability," you find pictures like Figure 4.17, showing one measured variable of the product following a gaussian distribution and its relationship with the tolerance interval, between the Lower Specification Limit (LSL) and the Upper Specification Limit (USL).

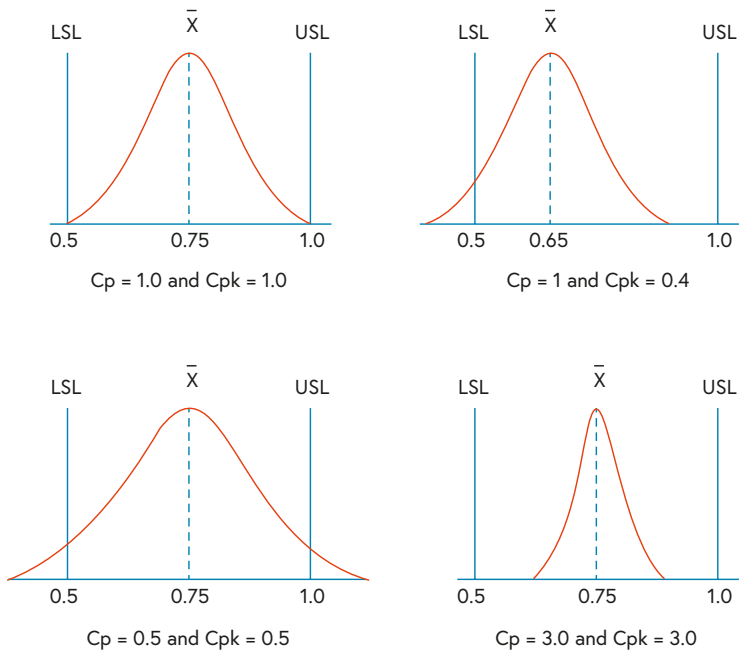


Figure 4.17 Process capability measurements

If m and s are the estimates of mean and standard deviation of the measurements, C_p and C_{p_k} are ratios defined as follows:

$$C_p = \frac{USL - LSL}{6 \times s}$$

$$C_{p_k} = min\left(\left[\frac{USL - m}{3 \times s}\right], \left[\frac{m - LSL}{3 \times s}\right]\right)$$

It applies, for example, to the example in Figure 4.18, where you cut a bar into rods that must be within 300 ±0.01 mm in length. The length of the tolerance interval is therefore 0.02 mm. You measure the lengths of actual bars with a sufficiently accurate instrument and compare its distribution with the tolerance interval.

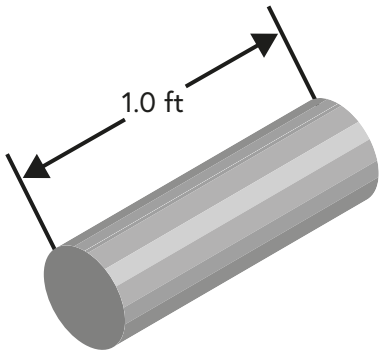


Figure 4.18 A simple product

In the classical analysis, you represent the precision of the process by the standard deviation of the lengths, and its accuracy by their mean, both of which you estimate from a representative sample.

If a rod is slightly too long, you can grind it down to length; if too short, it's scrap. The risks are asymmetric, and the process engineers are likely to aim for the upper end of the tolerance interval, where rework is possible. Let's say the mean length on a sample of 100 rods is 300.005 mm, and the standard deviation 0.0033. This makes the ±3σ limits 299.995 and 300.015 mm.

Because the ±3σ range is 0.02 mm, the C_p index that represents only precision is 1. On the other hand, the C_{p_k} index that takes into account both precision and accuracy is as follows:

$$\frac{(300.010 - 300.005)}{0.01} = 0.5$$

If the rod lengths can be modeled as the sum of a constant and a white noise, then $C_{p_k} \geq 1$ tells you that >99.7% of the rods are within the tolerance limits.

Because they are dimensionless ratios, the capability indices apply to any circumstance where you have one measured variable and a tolerance interval. In principle, it enables you to communicate process capabilities, whether about rod length or sugar content in cakes. This is the rationale behind the

reporting of process capabilities in terms of C_{pk} in many companies for processes that are more complex than cutting bars into rods.

Stop and Think 4.3
Which ones of the four processes in Figure 4.17 are capable?

Yield

Yield is an alternative and more general way of describing process capability. Even in the previous case of rod cutting, you can characterize it by the probability that a rod will meet tolerances, estimated on samples by the relative frequency of rods that actually do. In addition, not all observable quality characteristics are measured variables with tolerances. Many are, in fact, attributes, like the presence of scratches, bubbles, or runs on a painted surface, and any one of these defects makes it defective. It's a go/no-go situation, characterized by the probability of being defect-free, which is estimated by the yield of the process. If the yield is 80%, then this probability is also 80%.

The reasoning behind the C_p and C_{pk} above is not applicable here, as there is no tolerance interval. Control limits based on sample yields serve to validate improvements and detect accidental shifts. Their purpose is not to keep the yield at a fixed level, given that the best yield is always 100%. The common model for defectives in a sample is binomial, which has a single parameter, the probability of being defective.

Process capability in multiple dimensions

The C_p and C_{pk} do not generalize easily to cases with multiple variables. An automatic transmission case for cars has >2,000 critical dimensions. You can have a C_{pk} for each one but no formula to convert all these C_{pk}s to a single C_{pk} for the whole transmission case. In 2012, Santos-Fernandez and Scagliarini proposed a software package for the multidimensional capability index called MC_{pm} but it's not a simple combination of C_{pk}s for each dimension.[8]

Figure 4.19 shows the case of just 2 measured variables with positive correlation. For every unit produced, the tolerances tell you to keep the (V_1, V_2) values within the rectangle formed by the tolerance intervals $[V_1\ Lo, V_1\ Hi]$ and $[V_2\ Lo, V_2\ Hi]$. If (V_1, V_2) is a 2-dimensional gaussian, 99.7% of the actual (v_1,v_2) values will fall within an ellipse as shown and the 2D capability index is based on the ratio of the area of the rectangle to that of the ellipse. This generalizes to n variables $(V_1,..., V_n)$, with a tolerance hypercube and an ellipsoid.

As this looks like fitting an ellipsoid peg into a hypercubic hole, it begs the question of whether it makes sense to model the space of acceptable values of $(V_1,..., V_n)$ as a hypercube when, in fact, the space of values corresponding to good end products can be any shape, which electrical engineers call a *shmoo plot*.

8 Santos-Fernández, E. & Scagliarini, M. (2012). MPCI: An R package for computing multivariate process capability indices. *Journal of Statistical Software, 47*(7), 1–15.

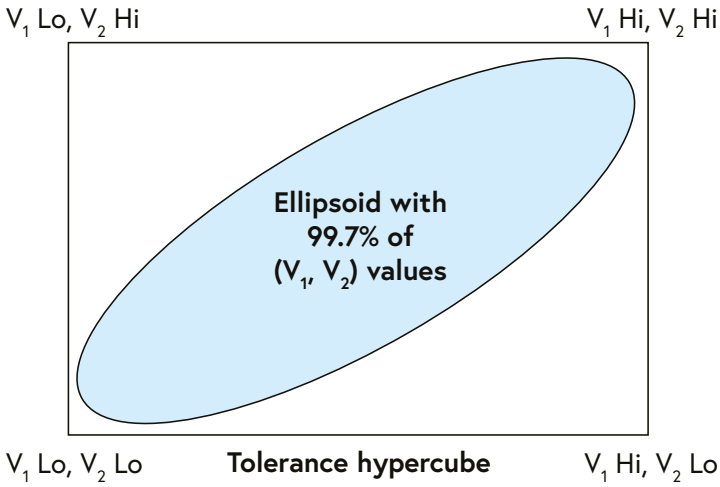


Figure 4.19 A 2-dimensional process capability plot

Yield rollups

Yields, on the other hand, scale up in many cases. Many organizations use yield as a high-level metric of process capability. If defects of different types occur independently, like paint scratches and transmission leaks, yields multiply. If 99% of a car line's production is free of paint scratches and 99.5% is free of transmission leaks, then 99%×99.5% = 98.5% is free of both.

Through a route with independent defects, the *Rolled Throughput Yield* is the product of the operation yields. Yield improvement at one operation will proportionately affect it, however, only if the defects are independent. Otherwise, an improvement at one operation may cause deterioration or an improvement at another as well.

On a route, the plain yield – the ratio of good to total units at final test – is a useful metric. It serves not only in production control but also in production line design. For example, if this yield is <95%, you cannot practically replace defectives with new starts, even with short lead-times. You overbuild systematically instead and maintain a buffer of finished units.

With a yield of 97% through a process segment of 5 to 10 operations, one-piece flow is feasible. Then, it will boost that yield further by rapid problem detection; with a yield under 90%, on the other hand, one-piece flow will be impractical.

The *First-Pass Yield* is another common metric but different production organizations may calculate it differently. It is the ratio of units that make it through the process without any rework to total units started. Varying interpretations of what counts as "rework" and pressure on managers to report high yields generate inconsistencies in yield calculations. If done consistently over time in a line, it is a useful tracker of improvements.

Interactions between characteristics

An integrated circuit, as opposed to a rod, is an open-faced sandwich with multiple layers. Each layer is a patterned thin film that interacts with other layers. Defects in the final circuit can be caused by complex interactions between different levels. All layers can be within tolerance limits and without defects, and the circuit may still not work. In this case, you do not improve overall yield by focusing on each operation. Instead, you apply data science to measurements and attributes collected on each unit, in-process, and at final test.[9]

4.5.3. **Measurement System Analysis**

You cannot take the accuracy and precision of measurement instruments or human observers for granted. It is well known to anyone who has "lost" *5* lbs from using a new bathroom scale, or observed differences of ± 0.5 lbs from stepping onto it multiple times. *Measurement System Analysis (MSA)* is capability analysis of the measurement or observation system rather than the process itself. As such, it is a *meta*-process capability analysis. When assessing process capability, measurement or observations errors compound with the variability of the process itself. Unless these errors are negligible with respect to the variation in the process itself, they can make the data unusable.

With measurements, the instrument adds its bias to the true value but fluctuations are not additive. If we represent the fluctuations due to the process by the standard deviation σ_p and those independently due to the instrument by σ_m, then the standard deviation for the total is not the sum of σ_p and σ_m but instead

$$\sigma = \sqrt{\sigma_p^2 + \sigma_m^2}$$

The bias of an instrument can be corrected by calibrating it or correcting the data. Then Gauge Repeatability & Reproducibility (R&R) uses analysis of variance (ANOVA) to separate the fluctuations between units with the same operator from the variations between operators.

Manual inspections for attributes are more subject to variations between operators than measurements. The details and clarity of the specs make a difference but there are factories with thick binders of specs for inspection procedures that do not contain the specifics of defects to watch out for. Measurement System Analysis (MSA) is a recently introduced term for these activities, and it encompasses Gauge R&R.

9 Senoner, J., Netland, T., & Feuerriegel, S. (2021). Using explainable artificial intelligence to improve process quality: Evidence from semiconductor manufacturing. *Management Science*. In press.

4.5.4. Process capability with Advanced Process Control

The C_p and C_{pk} are derived from Statistical Process Control (SPC), as developed in the 1920s when automatic controls were in their infancy. Control theory did not emerge as an engineering discipline until World War II. Sensors didn't proliferate inside production machinery until the 21st century. SPC focuses on characteristics observed on workpieces after completing operations, not on the control of processes in progress.

In 2006, Gary May and Costas Spanos[10] described the architecture in Figure 4.20 for process control in semiconductor fabrication:

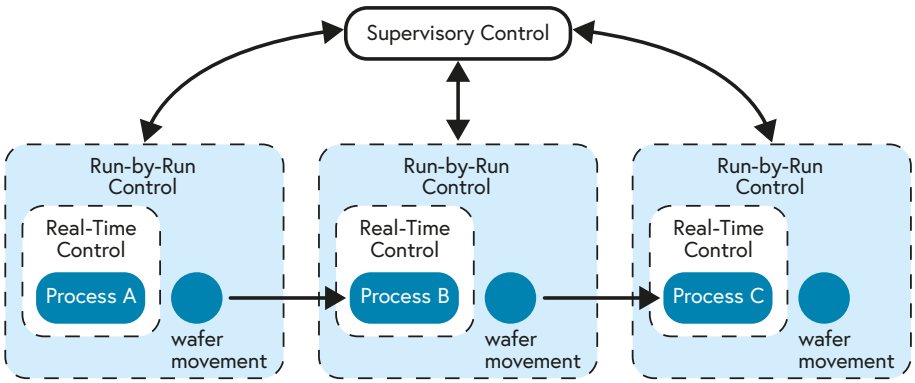


Figure 4.20 May & Spanos's Process Control Hierarchy

The different layers have the following functions:

- **Supervisory Control** ensures that each machine is running the right program for its current load. It also aggregates vital signs, and issues alarms. In May and Spanos's description, the Supervisory Controller compensates for variations observed at an operation by making adjustments to the next one.
- **Run-by-Run Control** replaces or enhances SPC. It performs multivariate analysis of the history of runs through the operation to tweak it for the next run.
- **Real-Time Control** encompasses all the automatic controls inside the machine. They range from mass-flow controllers to cyber-physical systems that simulate the process and tweak it as needed.

In 2022, suppliers of semiconductor production equipment offer Run-to-Run (R2R) control as part of Advanced Process Control (APC). R2R is another name for Run-by-Run. According to Costas Spanos, there are several such products on the market, and many of them evolved from early academic ideas and demonstrations. What makes them now possible is advances in IT and in-situ metrology.

Yields still work as measures of process capability in this context, where it is not clear how you would calculate C_p and C_{pk}.

10 May, G. S. & Spanos, C. J. (2006). *Fundamentals of semiconductor manufacturing and process control*. John Wiley & Sons.

4.6. Machining cells

Machining is a special form of fabrication, where the operations consist of removing material from the workpiece using cutting machine tools. It is often efficient to organize machining processes in cells, as exemplified in Figure 4.21.

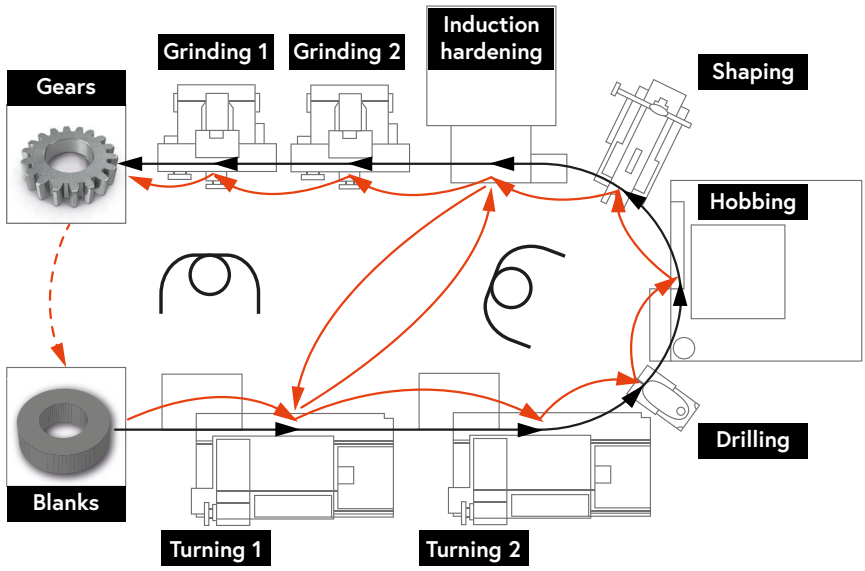


Figure 4.21 A machining cell

Machining cells are characterized by:

- **Materials flows.** Product units move through a sequence of steps without backtracking or iterations.
- **Where and how people work.** Production operators are surrounded by machines. The product units flow around the operator work area, without ever crossing it. Each operator usually attends to more than one machine, and these machines are of different types.
- **Product size.** Products made are usually small and light enough to be hand carried. In cases where some workpieces cannot be hand carried, all surfaces can usually be accessed from one side by rotating the part around the axis where it is longest. Workpieces that are large in 3 dimensions, like car bodies, need to be worked on from more than one side and are usually not processed in cells.
- **Number of operations.** Cell teams are preferably sized to experience the positive group dynamics of small groups. Less than three operators is too few for this to occur, but the labor needed to run the cell may not exceed that number. On the other side, more than 10 is too many. Even if it meets all the other requirements of a cell, a production unit with 50 operators is not a cell.
- **Number of operators.** The examples have no more than three operators working in a cell. There can be more, but that is rare at least in pure machining or fabrication cells. Where some assembly is involved, there can be as many as eight, or even ten, but a production line with 50 operators is definitely not a cell.
- **Station layout.** The machines are laid out so that the operator work stations of successive machines are as close as possible, and this design goal overrides almost every other consideration.
- **Operator station height.** All operator work stations are at the same height.
- **Density.** Machines in cells are much more closely packed than in the classical layouts. Even after allowing for maintenance access from the back of the machines, the conversion of a machine shop into cells typically frees up 30% of the floor space.

Unlike manual assembly, the work done by machines cannot be rebalanced at will between machines. You cannot use a lathe to hob a gear or a hobbing machine to drill holes. Consequently, it is inevitable that some machines are busier than others, and the cell's capacity is limited by that of its bottleneck. You are no more concerned about keeping the drill press of the cell 100% busy than you are about doing the same for the pallet exchanger of a machining center.

What some call "virtual cells," with workpieces flowing through physically separated machines, cannot be expected to yield the corresponding benefits. A real cell is a physical structure. The machines are close to one another and laid out according to the order of the process to facilitate the flow of materials and the movements of operators. There is no looping and no backtracking in the flow, and the physical location of a part inside a cell identifies unequivocally where it is in its process, so that, inside a cell, there is no need for tracking paperwork.

The cell in Figure 4.21 is U-shaped, which is the most common cell layout for the following reasons:

- **Pacing.** The same operator attends to the last and the first station, and is therefore able to match the pace at which blanks are released into the cell with the completion rate.
- **Visual control.** Each operator can have an understanding of the status of the cell from any point within it.
- **Staffing flexibility.** By making a large number of machines accessible to operators from a small work area, the U-shape makes it easier to modify the operators' task assignments in response to variations in demand.
- **Fixture flow.** While this is not the case in Figure 4.21, there are other processes such as printed-circuit board assembly in which workpieces travel between machines on fixtures. In this case, the U-shape allows the immediate reuse of a fixture when a work piece completes the process.
- **Quality.** The collocation of incoming blanks and completed workpieces makes the result of the work performed in the cell visible to its operators, and raises their quality consciousness.
- **Access control.** The U-shape's opening is the only one entry and exit point of the cell's operator work area, which allows the team to control access to its machines and tools.

Cells have clear operator work areas. The operator area inside the cell is free of any obstacle to the operator's movement: there are no chairs, no platforms, and no benches of any kind, and it is too narrow to hold shelves of parts. The operators' first impression typically is that it is too tight. However, once they realize how much easier the work is in a cell, they do not want to change back to the old way.

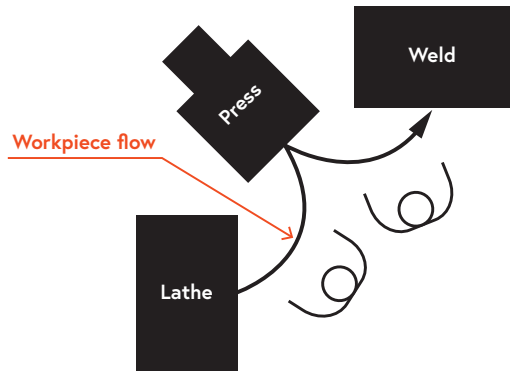


Figure 4.22 One piece flow beats basket-to-basket operations

Cells should be designed with one-piece flow. Initial cell implementations normally do not involve the automation of materials handling between machines within the cell. The most common pattern is for parts to be hand-carried between machines, and the major difference with what happens in a job-shop is that instead of having one operator pick a part in an input buffer and placing it in the output buffer, the parts move directly from one machine to the next with no buffer in between. The parts may even be loaded and unloaded in the same machine by two different operators, as shown in Figure 4.22.

While one-piece flow clearly minimizes the amount of WIP in the cell, the corresponding inventory reduction is not the main point. The main benefits of one-piece flow are the quality improvements that accrue due to the following:

1. One-piece flow minimizes the transit time of each workpiece through the sequence of operations in the cell. As a consequence, if any of the machines in the cell starts making defective parts, the operators detect it before the machine has had the time to do more damage.
2. As shown in Figure 4.23, by preserving process sequence, one-piece flow facilitates the tracing of problems. Not only does it make it easy to isolate the defective part but also to observe the parts made prior to the first defective part, to see, for example, whether the deterioration of the process was gradual or sudden.

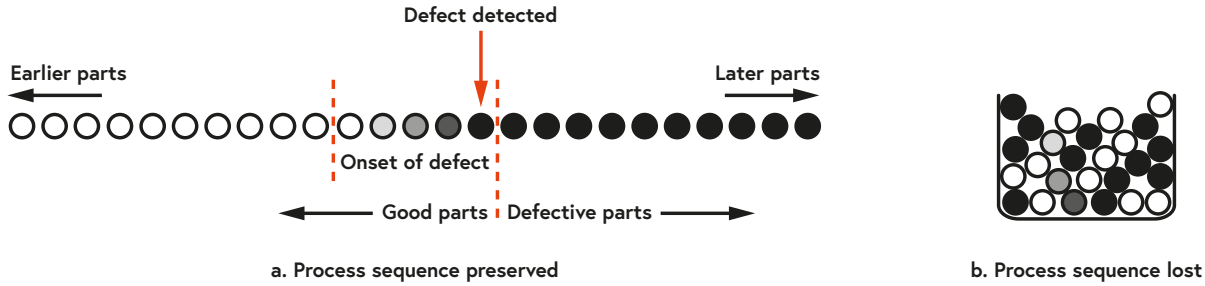


Figure 4.23 Preservation of process sequence in one piece flow

Some cells do not process individual parts but matched sets of parts going into the same end product. "One-set flow" is a legitimate variation on one-piece flow, applicable when the cell processes are stable enough for all the parts in the set, so that it is sufficiently rare that a set is unusable because of one defective part.

One-piece flow is not applicable with machines that process multiple identical parts at the same time when their entire capacity is required. While such machines are undesirable, they are sometimes inevitable, at least until engineers find an alternative that works one piece at a time. Often, these machines are "monuments" and cannot be circumvented nor replaced, but sometimes they are small enough to be integrated in cells. On a smaller scale, these machines are to cells as monuments are to the plant as a whole: they are rogue machines, and are not allowed to impose their mode of operation on the rest of the machines in the cell. While all the other machines work in one-piece flow mode, a full load accumulates in front of the rogue machine, and the downstream process works off a completed load downstream.

Box 4.5 **FIFO and LIFO**

FIFO is short for "first-in-first-out," and means that the parts that arrived at the station or into inventory first are used first. LIFO is short for "last-in-first-out," and means that the parts that arrived most recently are used first. While everyone agree that FIFO should be the preferred choice, many systems are accidentally designed for LIFO. Consider your cupboards at home: is your use of glasses and cups FIFO or LIFO? LIFO systems run the risk of parts getting dated, obsolete, or dirty. Even if it processes parts one at a time, a LIFO process still scrambles the message "written" on the parts by the process, making it more difficult to detect and solve quality problems.

4.6.1. The process of designing machining processes

Designing a manufacturing process requires answering questions in the right order. Choosing methods and sequences too quickly usually means going by what has been done on similar tasks before. The opportunities lost can be a reduction of 50% or more in process times. The purpose of a disciplined approach is to accelerate process development, not bog it down in bureaucracy. The resulting machining process is an effective use of the existing resources in the plant, or involves the acquisition of equipment with the lowest possible cost of ownership to do the job at hand.

From a technical standpoint, it involves process planning, equipment selection, layout planning and operator job design; from a managerial standpoint, teamwork among engineers, first-line managers, technicians and operators, the production and review of design documents, and project milestones. The process development sequence is not an alternative to concurrent engineering. It goes with it. Manufacturing input to product design is essential for production preparation to proceed quickly, and design changes may result from process development. The process development concepts remain valid but do not work as well with a design that is "thrown over the wall" to the manufacturing side.

While floor space is rarely the most expensive resource, it is sometimes treated in production preparation as if it were, often confining line layouts to oddly shaped, discontinuous areas where flow cannot be established. Machines are selected first, and process planning then becomes an exercise in figuring out how to make the needed operations on this machine. The selection of a machine is then made without the necessary inputs: the intended production volume over time and the definition of the process.

For machining cells, the following steps should be taken:

1. Plan for flow
2. Identify the takt time
3. Create the process storyboard
4. Technical implementation

STEP 1. PLAN FOR FLOW. In most machine shops, the latest, fastest and most flexible machines are used in priority over the older, slower, more conventional machines left over from previous generations. Engineers prefer to work on the most modern equipment, and managers are keen to show that the most expensive machines are being used, as opposed to sitting in overhead. This, however, is not necessarily what is best for the shop as a whole. Machines are often laid out based on space available, and neatly aligned with other machines along coolant drainage troughs, chip disposal systems or even power lines. Even around automatic machines, there are flows of work pieces, cutting tools, fixtures, and control programs that involve human intervention by machinists whose skills are key to the competitiveness of the machine shop, and whose effectiveness should be a key consideration in layout design.

The designer of a machining process must distinguish between primary and secondary operations. If the Maintenance department has the decision authority over machine selection, you may hear complaints from the Production department that the equipment has been chosen by people whose only interaction with it is "changing filters," as opposed to running parts through it. Conversely, if maintenance is not consulted, technicians may report afterwards that access to key parts of the machines is difficult or dangerous.

STEP 2. IDENTIFY THE TAKT TIME. In machining as in other processes, the takt time needs to be discussed first. The takt time depends on both the demand and the work calendar. The denominator in the takt time calculation is the demand for which the line needs to be designed. The anticipated demand pattern is an input from the marketing and sales side of the business. The decision of what rate to design the process for, however, is part of production preparation.

The following table shows an example of how the takt time varies from 10 min to 84 min when, to meet the same demand, the work time varies from 1 shift/day, 5 days/week in 1 line to 3 shifts/day, 7 days/week on 2 lines.

Table 4.3 Given constant demand, the work schedule sets the takt time.

Takt time	Shifts/day	Days/week	Lines
10 min	1	5	1
20 min	2	5	1
28 min	2	7	1
42 min	3	7	1
84 min	3	7	2

If the demand is sustained over time, it may make sense to give the plant the capacity to meet the peak demand. If demand peaks are seasonal and predictable in amplitude, as, for example, in construction equipment, then it may make sense to set the capacity below the highest peak and build ahead. If demand peaks come in random spikes, as in production equipment for semiconductors, a better strategy may be a second-source agreement with another machine shop. Whatever the pattern is, it is the first thing the team needs to discuss and resolve.

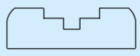
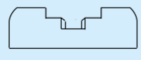
In addition, scalability also needs to be considered. The machining process is designed to support a given demand, but the actual demand on a particular day may be half that, and the process needs to run at half speed or operate half the time. When it runs at a reduced rate, the process should also use fewer resources. The initial discussion must also at least bracket the amplitude of fluctuations that needs to be supported.

STEP 3. CREATE THE PROCESS STORYBOARD The next step is to plan the processing steps for making the product. For this, a *process storyboard* is helpful. Table 4.4 shows one example. The original storyboard form this is based on, did not have "software" as a line item. Today, however, it is as much a part of the design of a machining operation as tools and fixtures. It may refer to a Numerical Control (NC) program in a Computer Numerical Control (CNC) machine, to the logic programmed in a Programmable Logic Controller (PLC), or to relay logic depending on the choice of process and equipment.

The sketch is not an engineering drawing but a crude rendition of a concept, drawn by hand with a thick marker pen. There are typically multiple options in the sequencing of cuts, each of which may need a separate storyboard. There are multiple options again in the choice of methods. A planer could be an alternative to face milling. Because the hole in the picture doesn't go all the way through, drilling and reaming cannot be combined in a single tool, but drilling and countersinking could be.

The process storyboard helps generate multiple options for the design of the process and the selection of the equipment to support it.

Table 4.4 The storyboard shows one methods to achieve the result

Part Number:	K-0123-206N			**Takt time:**	0:04:23	
Operations	**0**	**1**	**2**	**3**	**4**	**5**
Sketch						
Method		Face mill	Drill/ream	Counterbore	Countersink	Tap
Gauge		Snap gauge, CMM, spindle probe	Plug gauge	2 plug gauge	Taper plug gauge	Thread plug gauge
Cutting tools		End-mill, with or without inserts	Drill+Reamer	End-mill or counterbore with pilot	Countersink	Fluted spiral tap
Fixture/jig		Single unit versus multiple unit fixtures			Jigs	
Software		...				
Machine		Machining center 3-axis mill (vertical) 3-axis mill (horizontal)	Machining center Turret drill Gang drill	Machining center Turre drill Gang drill	Machining center Turre drill Gang drill	Machining center Turre drill Gang drill

STEP 4. REMAINING TECHNICAL STEPS The next steps include planning the details of secondary processes. One has to establish process capability and capacity, and then draft layouts, adding secondary operations like deburring and cleaning where needed as well as conveyance mechanisms to transfer parts between operations. The key elements to review should be the flows of materials and the movements of operators. This is true in other environments as well as in machining, but, in machining, there may be other materials flowing besides the work pieces, specifically the following:

- Cutting tools, from tool storage to mounting or rotating inserts, presetting, loading and unloading in the machines, resharpening, etc.
- Fixtures, from storage to pallets to machines and back.
- Chips from machines to recycling

Cells are an effective way to organize machining operations, but not a panacea, and there are many applications that are organized differently, particularly when using machining centers that process a part complete in one or more lays. Even where the primary process is not set up as a cell, however, the flows of work pieces, tools, and fixtures around the machines frequently afford the opportunity to set up "pseudo-cells," which have most, but not all, the characteristics of cells, and where many of the cell concepts can be applied.

The work combination charts described in Section 4.7 are used to design operator jobs. Their use is not limited to cells. They are useful wherever you have people interacting with machines, and the operation of CNC machining centers fits that bill.

4.6.2. Materials handling within the machining cell

If applicable, chutes can be used to convey parts between machines. On the other hands, bowl feeders or any other type of device that does not preserve sequence should be avoided. Pick-and-place robots are applicable, but only after the cell equipment has gone through all the first six steps of automation, and if it can be cost-justified in terms of number of operators removed from the cell.

When the size and weight of the parts make carrying by hand impractical, there are several possible options. Overhead bridge cranes are undesirable, because they are shared among multiple operations and their use must be coordinated with others. They are a part of the facility rather than the cell, and awkward to move. Jib cranes are local and can be placed just outside the cell between two machines, with their arm extending into the cell for operator access. When using cranes, the team must also think of how the parts are to be attached for transfer. When applicable, picking up a part with a magnet is most convenient. Using straps or clamps is much more time-consuming.

The disadvantage of cranes of any type is that they do not control the orientation of the part. It behooves the operator to manually keep the part from swinging about and to position it correctly in the next machine. Custom pushcarts, that hold the parts from below, offer better orientation control but they must either be usable as fixtures that dock into machines or include means of transferring the parts horizontally to and from the machines.

If this is a one-part cell, then the machines must be dedicated to this part, which means that the production volume must be sufficient to justify this. If the cell is for a family of parts, then they must have a common sequence of operations through these machines.

These issues play differently in machining and in assembly cells. In machining, there usually is one expensive bottleneck machine surrounded by auxiliary equipment whose utilization should not be a concern. If a $1M machine is underutilized as a result of being in a cell, it is much more of an issue than for a $5K drill press.

4.6.3. Additive manufacturing

Additive manufacturing (AM) differs from traditional machining and assembly processes. In AM, you build the part layer by layer, by depositing material from a nozzle in 3D printing or by routing a laser beam over a bed of metal balls. This happens in the build room of the AM machine. A digital 3D model of the product is a necessity to produce anything with AM. Therefore, AM is enabled by computer-aided-design (CAD) software or 3D object scanners.

It is essential to understand that AM is not one or two technologies, as illustrated with polymer and metal printing in Figure 4.24. There is a large variety of different AM technologies that work with different materials in different ways. The most common materials are different forms of plastics, or polymers, but today we can also print metal, ceramics, wood, and many other materials. We can even print biological mass and research promises to 3D-print organs in the future. Researchers at ETH Zurich, MIT, and elsewhere are developing "4D printing," which means printing a product that has the ability to change its properties after printing, for example shape or color. Today, it is also possible to print electronic circuits and biological tissue known as"bioprinting."

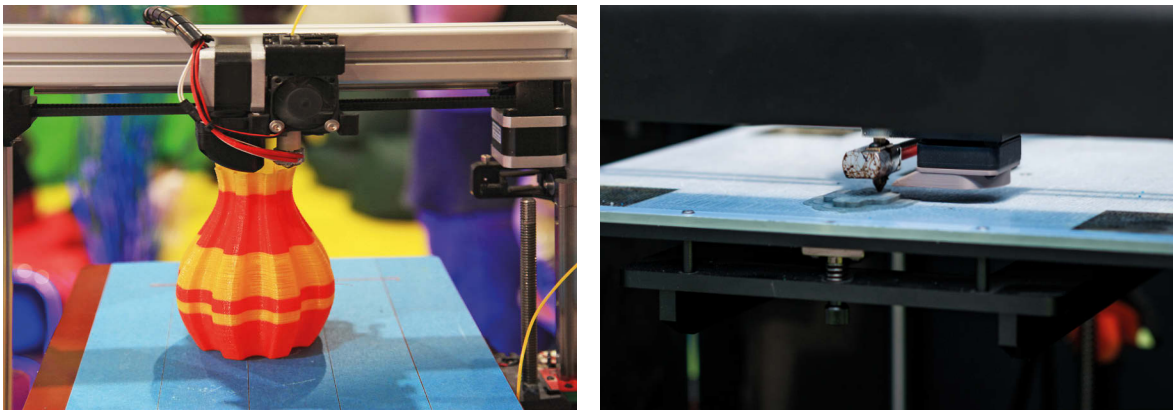


Figure 4.24 Additive manufacturing is much more than polymer (left) and metal (right) printing

AM can be a viable alternative to conventional manufacturing when all or some of the following conditions are fulfilled:

- Complex product geometries
- High degree of customization
- High demand variation
- Low product volumes
- High demand volatility
- Relatively small-sized products
- Medium to high product value

Successful AM applications usually involve a high degree of customization. In healthcare, for example, you can customize prosthetics, bone structures, orthodontics, hearing aids, and glass frames to individual needs. A large part of this industry has already converted. For example, all hearing aid implants in Switzerland are 3D-printed today. Box 4.6 discusses the perhaps most famous 3D-printed part in serial production: GE's fuel nozzle.

AM is also promising for spare part production, supporting after-sales services of all types of machines, electronic products, and furniture. The demand for spare parts is highly variable, in low volumes, and highly volatile. With AM, a service supplier can make spare parts on demand close to where they are needed, and when they are needed. A main challenge – besides the technical limitations of AM technology – is the preparation and management of the design of the product from development to end-of-life. Earlier CAD-drawings are not compatible with AM. A related hurdle is the increased risk of intellectual property theft and counterfeiting when everything is digitally available.

Note that manufacturing of tools such as jigs, fixtures, molds, and dies for traditional manufacturing processes share some of these characteristics and are therefore also good candidates for AM. All the above examples are small-sized products. The build room in 3D printers sets the limitations of how large the products can be. All examples also have medium-to-high value. Low-value products can usually be stored without a prohibitive cost penalty, making AM less lucrative for high-volume commodities.

Box 4.6 **Additive manufacturing of General Electric's fuel nozzle**

The famous 3D-printed GE fuel nozzle is the most cited industrial example (see Figure 4.25). In 2012, GE started piloting the metal printing of fuel nozzles. It reduced the bill of materials from 20 from several suppliers to one part manufactured in-house. Instead of a month lead-time, GE now can now produce it in a week. Of course, that has completely changed the established supply chains.



Figure 4.25 GE's 3D-printed fuel nozzle (Credit: GE)

The 3D-printed nozzle is also 25% lighter than its predecessor, and therefore saves fuel costs for the airlines. Furthermore, it operates in a challenging environment of high pressure and heat, and is still five times as durable as the conventionally manufactured nozzle.

Today, the 3D-printed nozzle is produced in high volume in the GE Auburn plant on 28 additive manufacturing machines. The Federal Aviation Administration cleared it for operation in 2015 and today it flies. By 2018, the Auburn plant had produced 30,000 fuel nozzles. In 2021, the number surpassed 100,000.

The GE fuel nozzle example shows that 3D printing is promising when you have parts with complex geometries. Traditional subtractive machining operations is difficult and expensive for such geometries. 3D printing can also provide product designs that are impossible with traditional approaches, such as complex internal structures. The solutions are lighter and more efficient for liquid or gas transport.

To date, AM remains a preferred technology for *rapid prototyping* and not so much for high volumes. Rapid prototyping is by far the most-used professional application of AM. For example, it accounts for about 90% of 3D printing applications in the automotive industry. However, for the industries with a good fit, change can come fast. As discussed, 3D printing has already disrupted supply chains in personal health industries and aerospace.

As AM reaches maturity and proves itself as a competitive technology, more and more processes will convert to it. While many companies invest in AM to produce prototypes or specific parts, some companies have made AM their core competency and sell it as contract manufacturers. Figure 4.26 is an example from the production floors at Norsk Titanium's AM factory in Plattsburgh, New York, USA. It is easy to see that AM processes are radically different from conventional processes.



Figure 4.26 A factory specializing in AM of titanium parts (Credit: Norsk Titanium)

When both volumes and quality requirements are high, 3D printing technology in 2021 is simply not economical or competitive. Industries that can use polymers have an advantage, but metal printing costs and speed must be slashed by a factor of 10 to 100 and quality must be greatly improved before it becomes competitive for the large majority of manufacturing applications.

> **Stop and Think 4.4**
> Look at the additive manufacturing factory in Figure 4.26. What do you find remarkably different from a traditional machine shop or assembly process?

4.7. Operator Job Design and work combos

In any shop where operators interact with machines, the design of operator jobs involves meshing the operators' own timelines with those of the machines in a cycle that fits within the takt time. The visualization of such a cycle requires what Edward Tufte[11] called a *narrative of space and time*.

11 Tufte, E. (1990). *Envisioning Information* (pp. 108–110). Graphics Press.

The most common form of space–time narrative in business is the *Gantt chart* from the early 1900s. Because the Gantt chart shows tasks as independent, free-floating bars, it doesn't show any relationships. As a result, while useful in *communicating* a schedule, it is not much help in generating one.

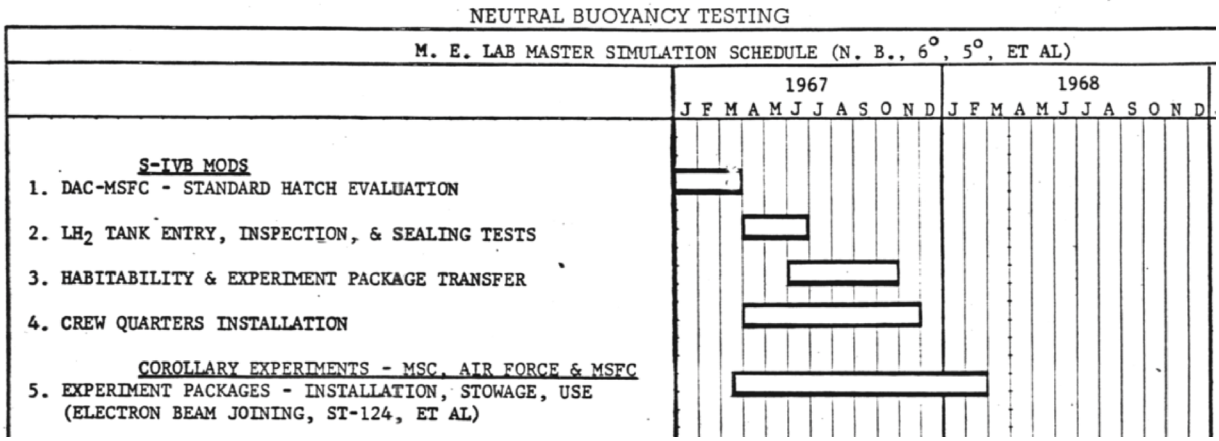


Figure 4.27 A Gantt chart from the 1960s

Tufte described the *Ybry chart*, that has been in use for 170 years in railroad scheduling and is a likely ancestor of the *work combination chart* from Toyota. The Ybry chart, by contrast, is focused on relationships like that of two trains running towards each other on the same track and able to cross only at locations that have sidings for this purpose.

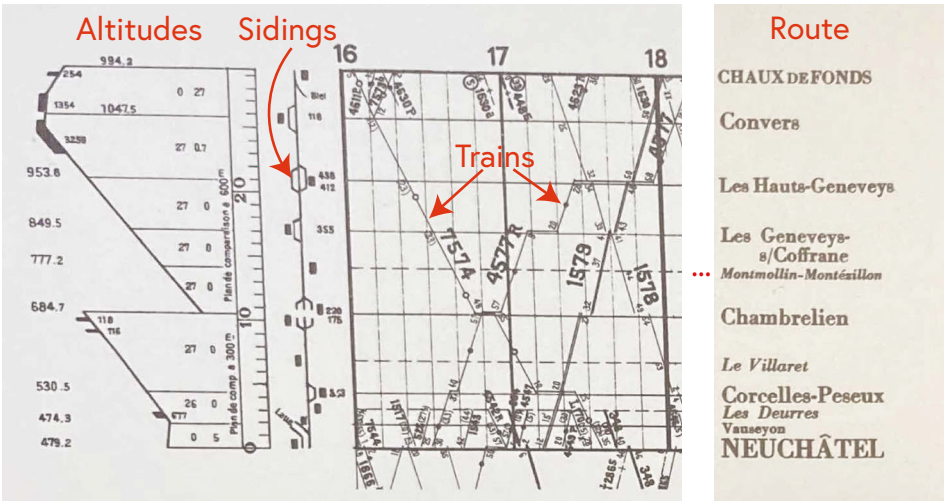
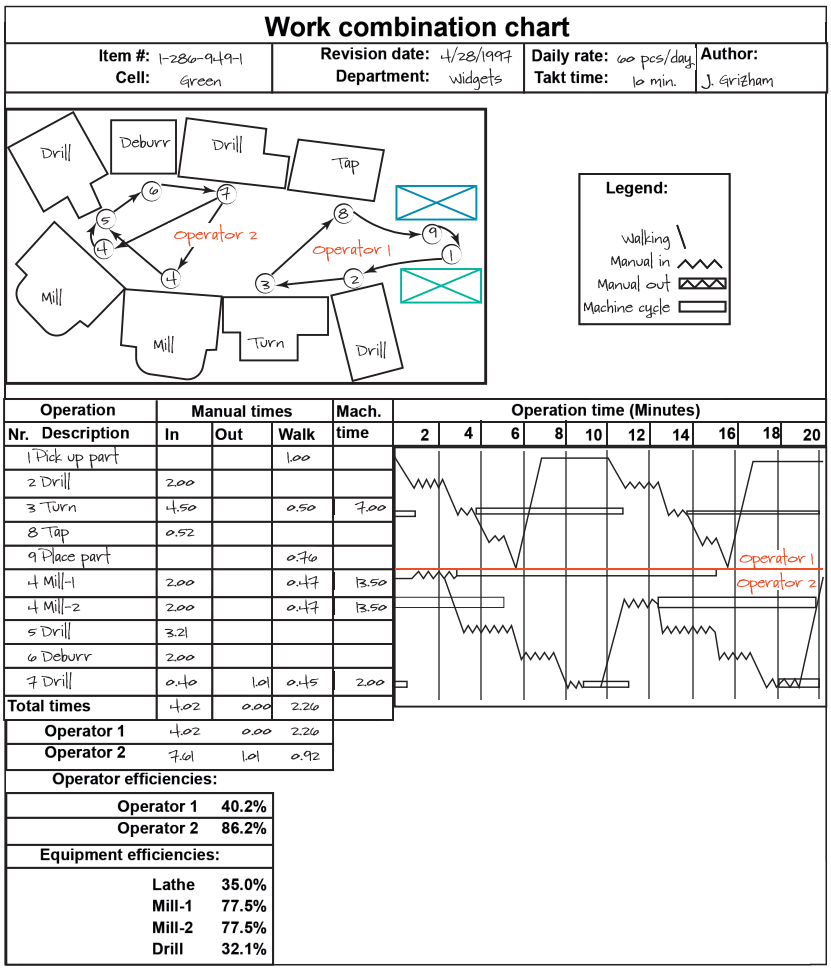


Figure 4.28 Ybry charts shows relationships

As shown in Figure 4.29, the work combination chart, or work combo, combines features of the Gantt and Ybry charts. It shows the time spent by operators walking between stations and starting automatic cycles during which the machines are unattended.

The example is a finished design, intended to be read by operators and posted on the floor. It is the end result of multiple iterations of design, involving video-based time studies and analysis conducted by

industrial engineers with participation of the operators. Operators generally do not *generate* these charts but they read and critique them.



Work combination chart

Item #: 1-286-9491 | Revision date: 4/28/1997 | Daily rate: 60 pcs/day | Author: J. Grizham
Cell: Green | Department: Widgets | Takt time: 10 min.

Operation Nr.	Description	Manual times In	Out	Walk	Mach. time
1	Pick up part			1.00	
2	Drill	2.00			
3	Turn	4.50		0.50	7.00
8	Tap	0.52			
9	Place part			0.76	
4	Mill-1	2.00		0.47	13.50
4	Mill-2	2.00		0.47	13.50
5	Drill	3.21			
6	Deburr	2.00			
7	Drill	0.40	1.01	0.45	2.00
Total times		4.02	0.00	2.26	
Operator 1		4.02	0.00	2.26	
Operator 2		7.61	1.01	0.92	

Operator efficiencies:	
Operator 1	40.2%
Operator 2	86.2%
Equipment efficiencies:	
Lathe	35.0%
Mill-1	77.5%
Mill-2	77.5%
Drill	32.1%

Figure 4.29 Work combo for two operators in a machining cell

In the TPS literature, these charts are called "Standard Work," while work instructions for individual tasks are called "work standards." The two are unfortunately often conflated when "Standard Work" is used to designate work standards and the original Standard Work is forgotten.

The alternative names proposed for it include "operator-machine balance" and "work combination chart." While it is the most descriptive, however, "work combination chart" is too long, which is why we shortened it to "work combo."

Work combos are not a common sight in factories outside of Japan. Generating them is a substantial effort. Cell designers who choose not to put it in, however, are leaving about half the potential productivity improvement untapped.

Summary

The most essentials points covered in this chapter were:

1. Processes are the building blocks of manufacturing.
2. There are several ways to classify manufacturing processes.
3. The process design must fit the characteristics of the product and its demand.
4. A manufacturing process for a new product can be designed following a seven-step approach.
5. Modern simulation software is of great assistance for process design.
6. In fabrication processes, the capability of the processes and the routing through stations are defining design criteria.
7. Process capability metrics are limited to one defining process quality parameter.
8. Additive manufacturing has the potential to radically change the process design of many products.

Test yourself

1. How do you think the processes used to manufacture this book can be described?
2. Use the ASME symbols to draw a process map of making a cup of tea.
3. Sketch the product-process matrix.
4. How would you design the assembly processes of an expensive Swiss watch?
5. How would you design the assembly processes of a mass-produced cheap watch?
6. What is "six sigma" process capability? What are the assumptions that are needed for the common process capability metrics C_p and C_{pk} to hold?
7. What is a process storyboard and when is it used?
8. What is a machining cell?
9. What is FIFO? How does it differ from LIFO?
10. In what manufacturing processes does additive manufacturing have the highest potential?

Further reading

Black, J. T. et al. (2008). *DeGarmo's materials and processes in manufacturing* (10th ed.). Wiley.

Frazelle, J. (2021). Out-of-this-world additive manufacturing. *Communications of the ACM, 64*(3), 58-62.

Harik, R. & Wuest, T. (2020). *Introduction to advanced manufacturing*. SAE.

Holweg, M., Davies, J., De Meyer, A., Lawson, B., & Schmenner, R. W. (2018). *Process theory: The principles of operations management*. Oxford University Press.

Hayes, R. H. & Wheelwright, S. C. (1984). *Restoring our competitive edge: Competing through manufacturing*. John Wiley & Sons, New York.

Hyer, N. & Wemmerlov, U. (2001). *Reorganizing the factory: Competing through cellular manufacturing*. CRC Press.

Madison, D. (2005). *Process mapping, process improvement and process management*.

Pyzdek, T. (2003). *Quality Engineering Handbook* (2nd ed.). CRC Press.

Swift, K. G. & Booker, J. D. (2013). *Manufacturing process selection handbook*. Butterworth-Heinemann.

Swink, M. L. (1998). A tutorial on implementing concurrent engineering in new product development programs. *Journal of Operations Management, 16*(1), 103–116.

5 Assembly

Assembly is commonly the final process in discrete manufacturing, and the most labor intensive. Therefore, the highest number of factory workers are usually found in assembly. To provide a sense of the scale of these types of manufacturing systems, the 2020–2021 *Occupation Outlook Handbook*, produced by the U.S. Bureau of Labor Statistics (www.bls.gov), estimates nearly 2 million workers in the U.S. alone are in fabrication and assembly, where the largest share of this number is assembly workers. The number is projected to decline by 5% from 2020 to 2030, but this projection was done before the political call for reshoring following pandemic-related supply chain shortages in 2020–2021.

The number of assembly workers in China is much higher but not readily available. To give a sense of the likely scale, the U.S. Bureau of Labor Statistic reported that 99 million people worked in the Chinese manufacturing industry in 2009. This number can be compared with the reportedly 14 million people working in US manufacturing in 2020. In fact, in *every* country, there are factories whose primary output is assembled products – from cigars to luxury watches, from cars to space rockets, from toys to smartphones. While "Made in China" can be found everywhere, we are surrounded by assembled products from all over the world.

Assembly is mostly human work, and many assembly processes – especially those with takt times below 1 minute – are mind-numbingly repetitive, dull, and sometimes dangerous. Automation can eliminate some tasks, but for many there still is no technology as flexible as a human. The challenge for manufacturers is to design and operate assembly factories that achieve high-performance by leveraging the abilities of their people and offering them interesting career paths.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Explain the general concept of assembly.
2. Decide when to use assembly cells.
3. Design jobs in cells.
4. Design assembly stations.
5. Organize the feeding of parts and materials to an assembly line.
6. Design disassembly processes.
7. Design remanufacturing processes.

DOI:10.4324/9781351110310-7

5.1. Introduction to assembly processes

Humans have been assembling since the first invention of a tool that could not be made out of one single piece. With the exception of a few forerunners, up to the 1st Industrial Revolution starting around 1760 in Britain, large products were assembled in a single location by multiple crafts persons and small products by a single crafts person assisted by a handful of apprentices in a small shop. Examples of large products assembled at that time are ships, railroad cars, or locomotives. Examples of small products are garments or shoes. The assembly factories of the 1st Industrial Revolution developing mostly from the mid-1800s brought together these crafts persons to work in parallel under the same roof.

This changed again in 1913 with the introduction of the *assembly line* at Ford's Highland Park factory. Historians of technology argue whether or not it was the world's *first* assembly line. For example, Mark Warren has found that the moving assembly line was implemented at EMF Motors in 1911 compared to 1913 for Ford. Coincidentally, the EMF plant was next door to the Ford plant. It is indisputable, however, that the Ford assembly line stands, worldwide, as the symbol for the *concept*.

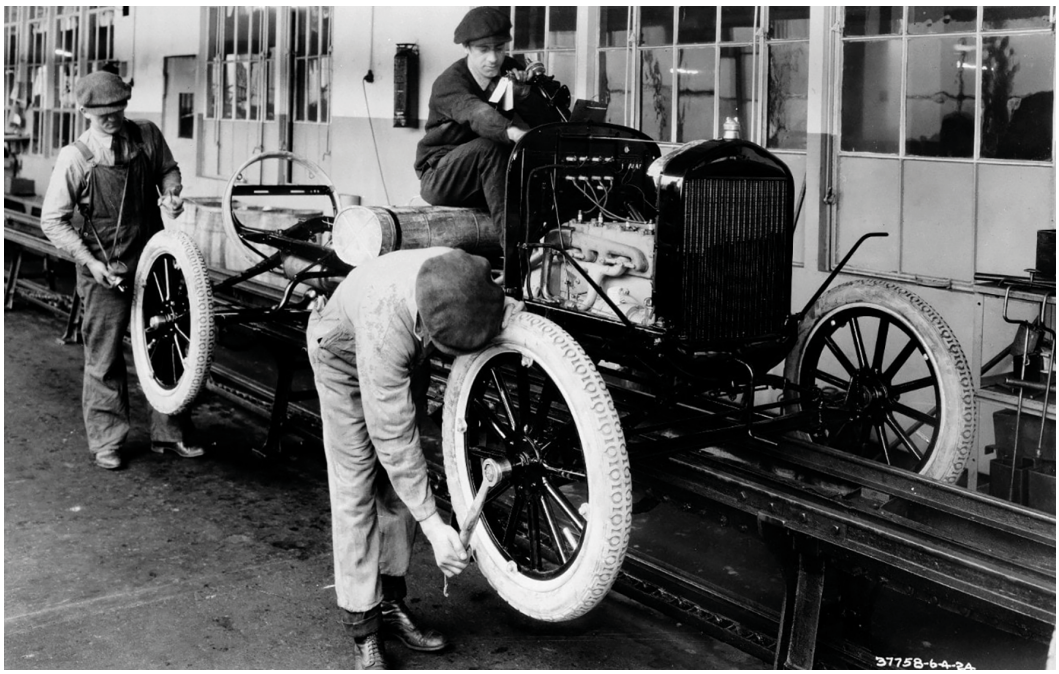


Figure 5.1 Assembly line at the Highland Park factory in 1924 (Credit: Ford Motor Company)

5.1.1. Assembly activities

> **Definition**:
> Assembly is any manufacturing operation with multiple, discrete items as input and a single one as output.

Casings and screws are discrete parts; joining two casings with screws is an assembly operation. Paint, on the other hand, is a bulk item and painting a workpiece is not considered assembly but *fabrication*. When an electronic component is soldered onto a board, the assembly operation is of the component and the

board; solder is used in the operation but it's a bulk item and is not listed as a component. The distinction matters because discrete components and bulk items are managed differently.

In manufacturing, assembly work is almost always repetitive. The only time it's not repetitive is in making custom products but, even then, the process is designed to leverage the commonality between the different units, for example by assembling a common semi-finished product and customizing it at the end. Repetitiveness allows assemblers to develop muscle memory and go "on automatic," which means executing a manual assembly operation accurately without thinking about it, while still being able to notice anomalies. Repetitiveness, however, also makes assembly work dull, particularly when the process cycles are short.

Assembling parts is predominantly characterized by *joining*, which is therefore considered as the primary activity. However, further secondary up- or downstream activities – like handling, controlling, and adjusting – are also necessary. Assembly activities can be distinguished as follows:

- **Joining**: screwing, filling, pressing on/in, riveting, joining through reshaping, welding, gluing together, soldering, textile joining, etc.
- **Handling**: storing, moving, securing, etc.
- **Controlling**: checking, testing, measuring, etc.
- **Adjusting**: through reshaping, separating, joining, setting, after treating, etc.
- **Secondary activities**: marking, heating, cooling, cleaning, deburring, imprinting, covering, withdrawing, unwrapping, lubricating, spraying, sealing, etc.

Assembly activities can also be separated into multiple levels of *subassembly* and *final assembly*. A subassembly is a unit assembled separately but designed to be incorporated with other units into a larger manufactured product. A subassembly may go directly into a finished good or into higher-lever subassemblies.

The product structure is called a bill of materials (BOM). In a well-designed assembly process, the same components and subassemblies are used in multiple higher-level subassemblies and finished goods, as shown in Figure 5.2. The structure shows different assembly stages. Each box is either a purchased component or a subassembly, and the arrows are "goes-into" relationships, with quantities and lead times attached. With an increasing number of items, the complexity of the product increases, and usually so does the number of subassembly levels.

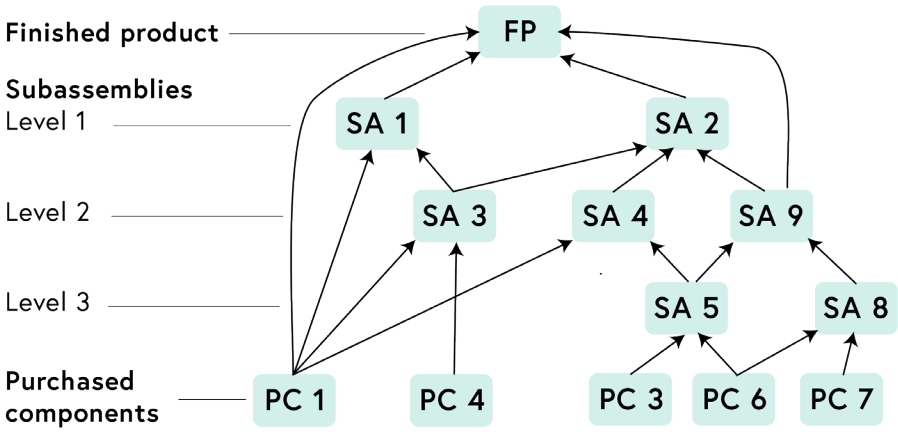


Figure 5.2 Structure of a product

5.1.2. The role of assembly processes

Assembly is the keystone of the overall production system for discrete manufacturing. It is the last phase before the product can ship, but it is also often the first step after which you can test its function. The assembly process is therefore not only responsible for finishing products on time, efficiently, and in the volumes required but it is also accountable for *initial quality*. Initial quality concerns defects reported by customers who have just bought the products, and these defects can often be traced back to manufacturing. By contrast, defects that manifest themselves through high maintenance costs over years are usually issues with the product design.

The ease of assembly and the quality of the resulting products is dependent on design decisions made for the upstream processes. A product that has not been *designed for assembly* may have more types of fasteners than necessary, requiring different tools. It may also cobble together subassemblies that do not work well together. Incomplete work plans are ultimately detected during assembly and need to be fixed then. Assembly is therefore sometimes an "error collection bin" for the upstream processes. In order to fix these errors, a high degree of flexibility is required in assembly.

Humans play a decisive role in making assembly flexible. Although automation is progressing, assembly remains a more heavily human process than fabrication. Humans are particularly difficult to replace in processes that require a response to errors and anomalies. Therefore, the design of assembly processes must be centered on operators and consider the mix of sizes, ages, and gender in the population from which they are drawn.



Figure 5.3 Assembly of Airbus A320 (Credit: Airbus)

5.1.3. Assembly performance factors

The two main factors in assembly performance are *part supply* and *assembly work design*.

Part supply is the most important factor in assembly productivity. "If only we had all the parts we need, we could assemble everything without any problems!" is a typical lament of assemblers, supervisors, and managers. Working with logistics and purchasing functions is clearly essential in running assembly operations. But assembly organizations cannot focus all their attention on suppliers; *in-plant logistics*, the last link of the supply chain, is under their control. In-plant logistics is the way the part physically gets from the unloading dock into the assembler's hands. The assembler may grab the part by extending his arm or by walking 20 steps to a shelf. It is the internal organization of the assembly area that sets which it is.

The second major factor in assembly productivity is the design of the assembly work, including:

- The subassembly structure.
- The sequencing of assembly steps.
- Their allocation among assembly stations.
- The sizing, fixturing, and tooling of assembly stations.
- The design of assembly procedures.
- The means of controlling the quality of assembly work.

5.2. Assembly lines

Although the assembly line is often synonymous with assembly operations, it is only one way of organizing work in assembly factories. Since it was invented in the early 1900s, the assembly line is hailed for its productivity. All the essential elements that make up an assembly line workstation are illustrated in Figure 5.4.

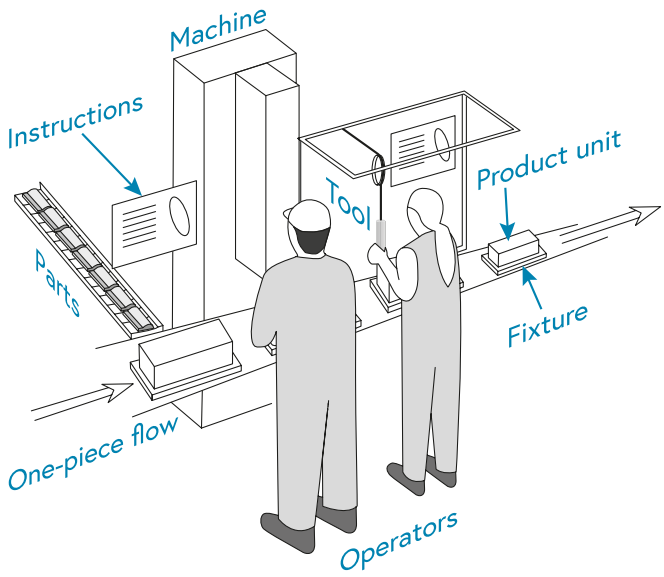


Figure 5.4 A workstation at the assembly line

The following are key points to keep in mind:

- The flow of product units down the line should be one piece at a time, not one batch at a time.
- Flowing one single-unit *kit* at a time, however, is consistent with the definition, because all the parts in it eventually go into the same product unit.
- The flow of the line is clearly defined but the direction of work flow at each station can go both ways: it can be left to right or right to left, or both if work takes place on both sides of the line, as is the case, for example, in the final assembly of cars.
- Backtracking should not be allowed. Product units should not return to stations they have already visited as it hurts flow and productivity.
- Some elements, like instructions, are mandatory at every station. Others, like machines or even fixtures, may or may not be needed depending on the situation.

Box 5.1 **Counterclockwise versus clockwise flow**

Some lean consultants insist that the flow of work in U-shaped cells should be counter-clockwise, or right-to-left and claim it is more productive than left-to-right. They argue that it should be so because most people are right-handed or because "all lathes have the head-stock on the left." Even if most lathes do have the headstock on the left, there is no reason why lathes should dictate the direction of flow for manufacturing operations that don't use lathes. As evident in Figure 5.5, car assembly lines have both left-to-right and right-to-left flows working at the same takt times.

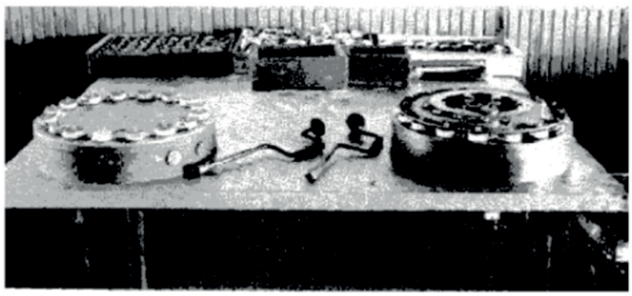


Figure 5.5 Car assembly lines assemble on both sides, hence left-to-right and right-to-left (Credit: Audi)

5.2.1. History of the assembly line

While it is clearly not simple to implement, the assembly line concept itself is straightforward. Why then did we have to wait until 1913, a good 150 years into the industrial revolution, before anybody tried it? It is not as if high-volume assembly didn't take place before then: steam engines had already been made by the thousands, railroad cars by the tens of thousands, and sewing machines by the millions. The dominant method back then was bench assembly, where the product unit does not move, all the equipment and parts are brought to it, usually in kits, and where one assembler or a team of assemblers moves around it.

In 1913, Ford engineers experimented with the concept of the assembly line for the Model T magneto in Highland Park, MI, as shown in Figure 5.6. On the first day, the labor time required to assemble a magneto – an electrical generator for ignition no longer used in cars – went from 20 minutes to 13 minutes. Within 8 months, it was down to 4 minutes, and it is this success that motivated Ford to apply the concept throughout the plant.



Bench assembly of the magneto



Ford's first assembly line

Figure 5.6 Ford Model T magneto assembly before and after the assembly line (Credit: Ford Motor Company)

Stop and Think 5.1
Compare the bench assembly and assembly line in Figure 5.6, what are the key differences in terms of work tasks and logistics?

Figure 5.7 compares bench assembly to line assembly. *Division of labor* is usually the first advantage of assembly lines that comes to mind, but there are many other, equally compelling benefits. Each component needs to be delivered only to the few stations where it is used, as opposed to all stations. The product itself moves with the line. Each station only does one operation and does not need to be set up for every operation.

Note that if division of labor were the only advantage of assembly lines, the concept would not be beneficial at low volume. The other advantages, however, remain at low volume. If we make the same product at a takt time of 100 minutes, we only need one operator, on a bench or an assembly line. The bench will be set up many times. In the assembly line, the operator moves the product on a pushcart between stations that are permanently set up with the fixtures, tooling, instructions, and parts needed for one step. In fact, the only area in which the bench wins over the assembly line is job satisfaction, but also that advantage can erode if the bench operator sees the efficiency of the line.

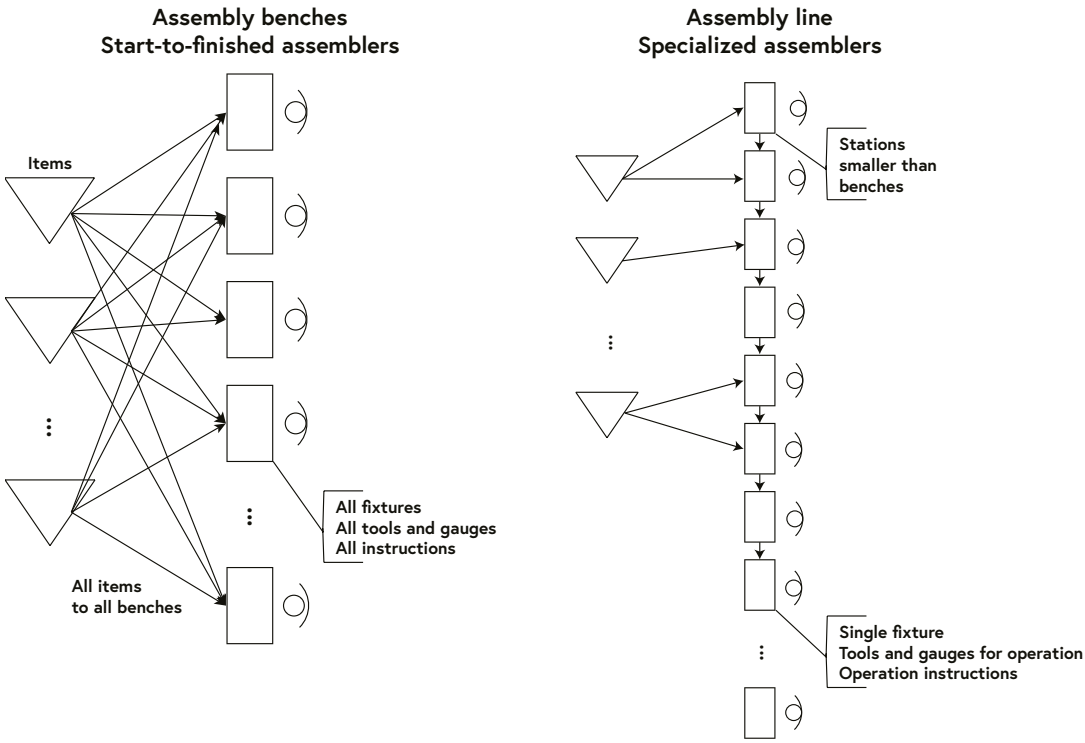


Figure 5.7 High volume bench assembly versus assembly line assembly

5.2.2. Where the bench is better than the line

In both high and low-volume production, the assembly line almost always outperforms the bench, yet there are exceptional circumstances that negate the advantages of assembly lines:

- The degenerate case where **the assembly process is very simple**, so that the assembly line only would have one station, and is indistinguishable from a bench. This happens, for example, in low volume assembly of simple hand tools, with fewer than 10 parts.
- **The product unit is so small** that everything that goes into the product fits on a bench and the assembly of it is a complex process that requires "human touch." Examples include high-end jewelry, mechanical door locks, and watches. If a human touch were not needed and volumes were large, automation would be superior to human assembly.
- **The product unit is too large to move.** While it is made of sections that move, a tanker hull or a offshore oil rig is assembled in one dry dock.
- **The unit is a prototype or a one-of-a-kind** system with no process similarity with anything else the company makes. This is the residual activity that remains after P-Q analysis has separated out repetitive production. Not only is there no takt time, but the assembly process itself has to be custom. Planning mistakes are common, resulting in components being disassembled and reassembled several times.

5.2.3. **Human work at assembly lines**

It is common to think of assembly line work as dull and hard work. It is a valid concern. Human factors have been the weakness of the assembly line since its invention. Ford quickly experienced a dilemma that has affected assembly lines ever since. On the one hand, assembly lines beat bench assembly on both productivity and quality. On the other hand, assemblers universally prefer bench assembly. The famous factory scene from Charlie Chaplin's movie *Modern Times* from 1936 mocks the repetitive work at the assembly line.



Figure 5.8 Publicity photo of Charlie Chaplin for the film *Modern Times* (1936)

At Ford, operators responded to the introduction of assembly lines by leaving the company. Turnover became such a problem that Ford introduced the "$5-day," doubling the market wages of the time as an inducement for workers to stay. Even as, over the decades, the $5-day turned into the $20-hour, this dilemma has not been resolved to this day. Even operators who "like putting things together" often dislike doing it on an assembly line. They accept the work because they cannot find or do not search for another job with comparable pay at their level of skills near their homes.

Over time, most assembly workers get used to the work or decouple it from their free time by accepting it in return for economic reasons. Often they find solace in the camaraderie of the shop floor and price in the product. While there are ways to make the work safer, more secure, and less boring, no techniques have been invented yet that make assembly on a line the sort of work children dream of doing when they grow up. It is, however, honest work that can earn a living.

While paying assemblers twice as much to be four times more productive is affordable, it should not be viewed as a permanent solution. Fundamentally, assembly line work should be automated away to the extent possible and feasible. This is, however, not easy. Recently, both Tesla and Mercedes Benz have removed automation from the assembly line to become more flexible. What concerns us today is what to do in the meantime. Most people in manufacturing work on an assembly line. Today and for decades to come, this work will still largely be done by people. It is the responsibility of manufacturing managers to make the work as safe and interesting as possible, and pay a fair salary for this type of work.

Box 5.2 **Volvo experiments in the Kalmar and Uddevalla plants**

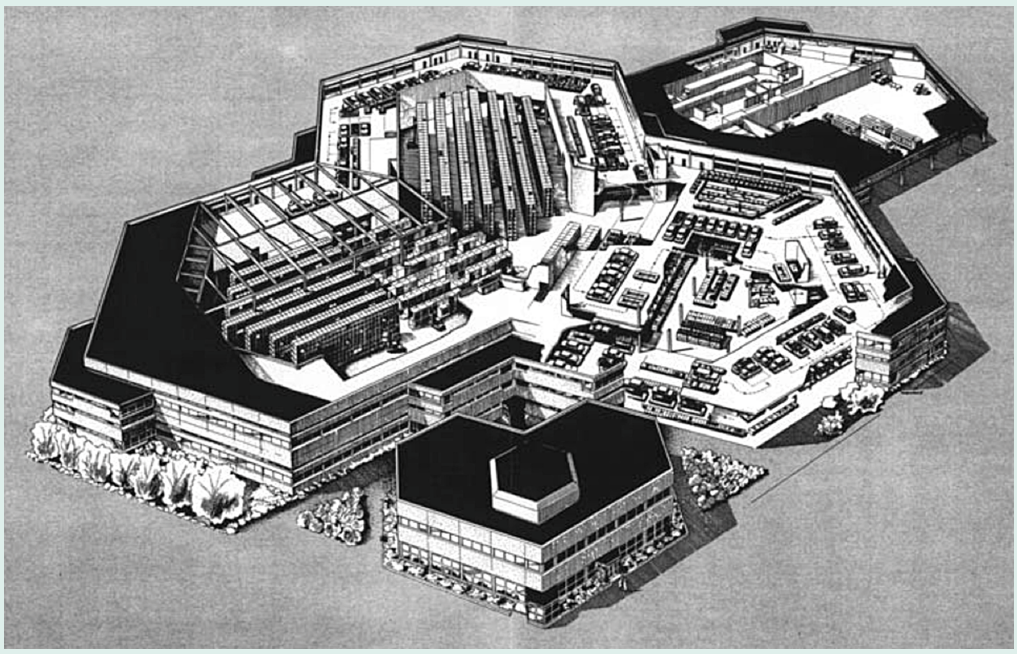


Figure 5.9 The Kalmar factory in 1974 (Credit: Volvo)

Following the Scandinavian tradition of workplace democracy, worker participation, and flat organizational hierarchies, Volvo developed and deployed a new way of thinking in the automobile industry between 1960 and 1994. The famous experiments in the Kalmar and Uddevalla factories tried to abandon the production line for assembling cars by introducing a more humane way: "dock assembly." In dock assembly, a team of workers builds larger parts of the car and travels with the dock as it moves through stations in the factory. It is characterized by a cell-based shop floor organization where production teams were given full responsibility for more complete assembly tasks. Figure 5.9 shows the innovative Kalmar factory and its iconic beehive layout and Figure 5.10 shows dock assembly.



Figure 5.10 Dock assembly in the Kalmar factory (Credit: Volvo)

Despite the short-term positive effects, the Uddevalla and Kalmar plants were closed down in 1993 and 1994 respectively, and the Volvo experiments were generally judged as failures.

"Volvoism" was only to a limited extent adopted in the European automotive industry during the 1990s through the diffusion of team-based production. Even if the experiments in the end proved non-competitive on a world scale, Volvo believes that the experiments were worth the effort.

5.3. Assembly line designs

Not all assembly lines are straight. In fact, long straight lines are often inefficient. The next paragraphs introduce different types of lines.

5.3.1. All assembly work done in one single line

For manufacturing, the simplest design has no subassemblies at all. All the work is done in one sequence, and all that needs to be done to coordinate and pace the work is to make sure that all product units proceed to the next station at takt intervals, as shown in Figure 5.11.

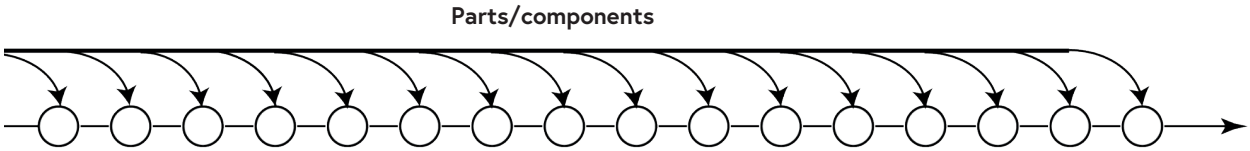


Figure 5.11 Assembly in one single line

5.3.2. Final assembly line with subassembly feeder lines

Within that single sequence of operations, however, you find subsequences of steps affecting only one small part of the product while the rest of the product waits, serving only as a fixture. In addition, the product may be an inadequate fixture, providing inconvenient access to the parts being worked on. It is typically much larger than a dedicated fixture would be, and it is not designed to orient this subset in a way that is convenient for assembly.

Where we observe these things, it is a good idea to think about making that subset a subassembly and put it together on a feeder line with a fixture designed specifically for it. This allows work to proceed in parallel on the subassembly and on the rest of the product, as shown in Figure 5.12.

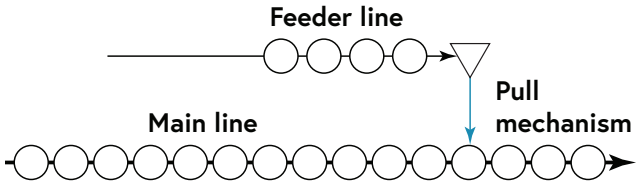


Figure 5.12 Subassembly feeder line

Dashboard assembly in cars is a case in point. Traditionally, it was done inside the car body by an assembler working in a succession of uncomfortable positions. The manufacturer pays for this, however, by having to coordinate the feeder line and the main line. Typically, the feeder line has an output buffer, from which subassemblies are pulled for use on the main line. As feeder lines proliferate, this coordination problem grows.



Figure 5.13 Honda operator, in the company's iconic white coverall, installs front-end module on feeder line, Indiana USA (Credit: American Honda Motor Co.)

Modular assembly

Auto industry executive José Ignacio López promoted the concept of "modular assembly" (Figure 5.14) at GM and Volkswagen. This means having suppliers put together major subsystems of cars and cutting down the amount of work in final assembly by about 90%.

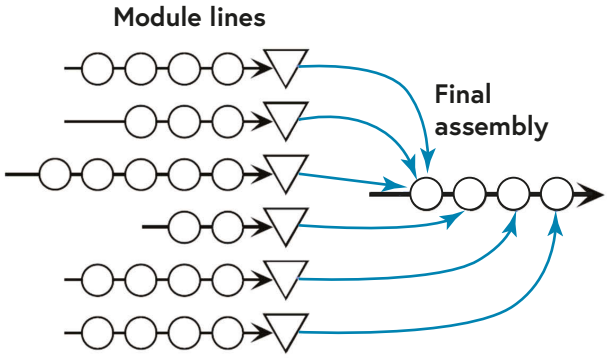


Figure 5.14 Modular assembly

Modular assembly is a common practice in electronics manufacturing. Final assembly of PCs, for example, is simple because most of the complexity is in such modules as motherboards, processor chips, and disk drives, which are assembled by suppliers.

5.3.3. Type of assembly line by product groups

Assembly lines can also be differentiated based on the diversity of products they can handle. There are three types:

- **Dedicated lines** make one product, possibly quite complex, for months or years, during which volume is adjusted up and down numerous times. Whenever major rate changes occur, groups of engineers spend weeks or months "rebalancing the line," which means calculating how many operators need to be added or pulled out.
- **Mixed-flow lines** alternate between products one piece at a time. In addition to the above, this requires engineering the assembly times to be as close as possible for the different products, which is done by doing more subassembly work or more automation on some products than others.
- **Batch assembly lines** alternate between products one batch at a time. This is much more common than true mixed-flow lines due to the widespread belief among assembly managers that it is more efficient. If at all true, this is only if you fail to consider the implications on the flows of materials, particularly when the mix fluctuates or when quality problems occur.

5.3.4. Assembly line design in automotive factories

For car assembly, straight assembly lines is a thing of the past. Except for modular assembly, final assembly lines for cars are so long and so large that real estate becomes a consideration. A straight line would be over half a mile long (~800 meters) but no more than a few hundred feet wide (~50 meters), which are awkward dimensions. The delivery of parts to the line, however, is a more important concern. The transportation aisles in straight assembly lines are usually inconveniently long. If you carry parts for delivery all the way along, you need a long train. If you have only one straight line, then you can have receiving docks all the way along, to bring parts straight from trucks to the line, but then it means many docks, spread far apart. Modern car assembly lines snake around as shown in Figure 5.15.

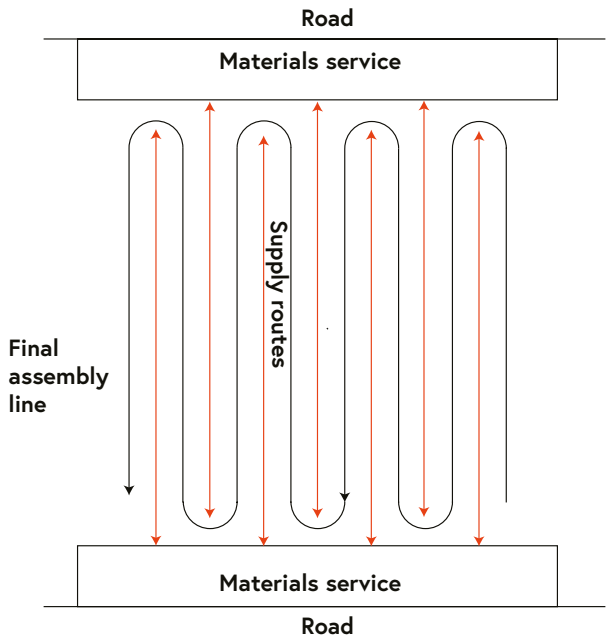


Figure 5.15 Snaking assembly line

Why do automobile companies adopt this shape? Some argue that it allows operators to communicate more easily across sections of the assembly line. Figure 5.15 highlights what we believe is a more compelling reason to adopt this type of design. Parts can be brought in through two compact materials service areas on opposite sides of the line, with parts supply shuttles going into the folds of the snake and being able to supply immediately two linear feet of assembly line for each linear foot of transportation aisle.

While this is a discussion of car assembly, its purpose is not to hold that industry up as a model but to point out the issues that drive this type of decision. Assembly lines for airplanes, electronics, or toys have different shapes, based on volume, product shape, or part structure, for instance, and on the type of analysis we discussed here in the special case of car assembly.

If we start running the snaking assembly line, it quickly becomes apparent that its operation is not as simple as Figure 5.15 might suggest. In the folds of the snake, only one side of each station is accessible from one of the materials service areas. When assembly work proceeds on both sides of a product, however, there frequently are mating left and right-hand parts assembled on both sides at the same station, and coming from the same supplier. In this layout, then, trucks from the same supplier would have to deliver the left-hand part in the South area and the right-hand part in the North area, and both service areas would have to coordinate deliveries to the line to ensure that balanced quantities of the left and right-hand parts are available.

A solution to this problem is shown in Figure 5.16 elevating the product for transfer between line segments opens access to both sides of all stations from both materials service areas.

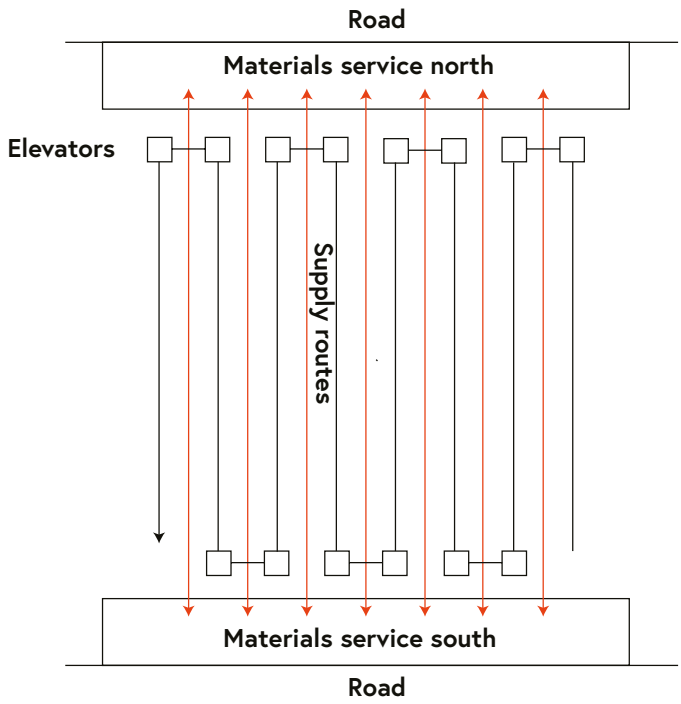


Figure 5.16 Snaking line with elevators at the end of each section

This in turn raises the question of regulating traffic in transportation aisles that serve to deliver parts to the same locations from opposite ends and still need the ability to deliver to both sides of the aisle in order to retain the advantage of delivering to two linear feet of line from each linear foot of transportation aisle.

In Germany, some of the leading car companies have been making yet another attempt to return to the job-shop. They have speculated that assembly lines are coming to an end because of their limited flexibility. They are betting that computer controls can enable the traditional job-shop to introduce and customize products faster than an assembly line. Instead of products moving down a fixed production line at a certain takt, the idea is that algorithms can allocate free capacity to production orders in real time. Figure 5.17 illustrates the concept.

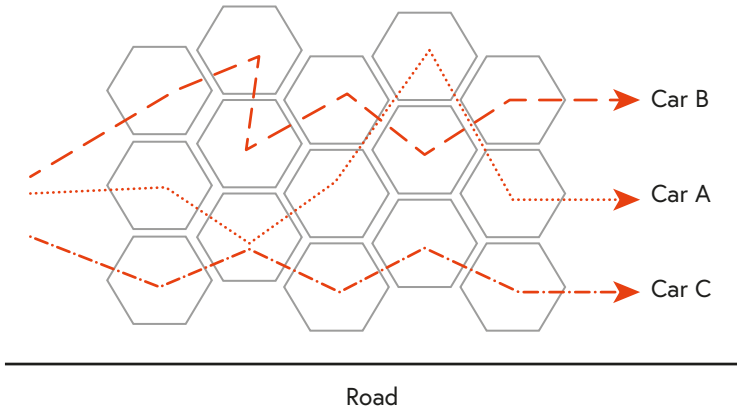


Figure 5.17 Vision for flexible, reconfigurable assembly lines in the automobile factory of the future

Electric cars are seen as a potential business case for implementing this idea, partly because they require only a third of the components of combustion engine cars. While the first experiments in Germany were abandoned, Audi is said to have implemented a pilot in its assembly factory in Győr, Hungary. Interestingly, Tesla chose a classic assembly line design for its assembly operations of all-electrical vehicles in their "Gigafactory" in Berlin-Brandenburg, Germany, which opened in 2022. Renault also kept the classic assembly line for its Renault Mégane E-Tech production in Douai, France, but has implemented a logistics system with 100% kitting of parts delivered to the assembly line on AGVs.

Stop and Think 5.2
Consider the line layout in Figure 5.17, what do you think are some key challenges related to job tasks and logistics?

5.4. **Assembly cells**

An assembly cell is a special case of an assembly line. It that can be operated by a team of maximum 8 to 10 operators. An assembly cell provides all the benefits of an assembly line, and adds those of being run by a cohesive team of multifunctional operators, which not only improves job satisfaction and morale, but also raises quality consciousness. The team controls the release of work to the first step in the sequence, and is responsible for putting out product units at the required pace. Most cells are U-shaped, with product units flowing *around* a long and narrow operator work area. This pattern supports flexible staffing, and the number of operators can easily be adjusted up or down with changes in demand.

In assembly cells, the operators work inside the U-shape on parts fed from the outside. While this is a desirable mode of operation, there are many assembly processes that do not fit this pattern, the best known being the final assembly of cars, which takes on the order of 200 stations, and engine assembly, which takes about 50 stations.

U-shaped cells work for products that are small enough in at least two dimensions to support feeding parts across the assembly stations and do not require simultaneous assembly on two or more sides. Most manufactured goods meet these conditions, but there are notable exceptions, such as cars, large household appliances, production machinery, or airplanes. Where U-shaped cells do not suit, it is often possible to reap most of their benefits through "pseudo U-shaped cells," where the material flow pattern is different, but we still have the operator work area inside and the operator team concept.

Figure 5.18 shows the layout of a typical U-shaped assembly.

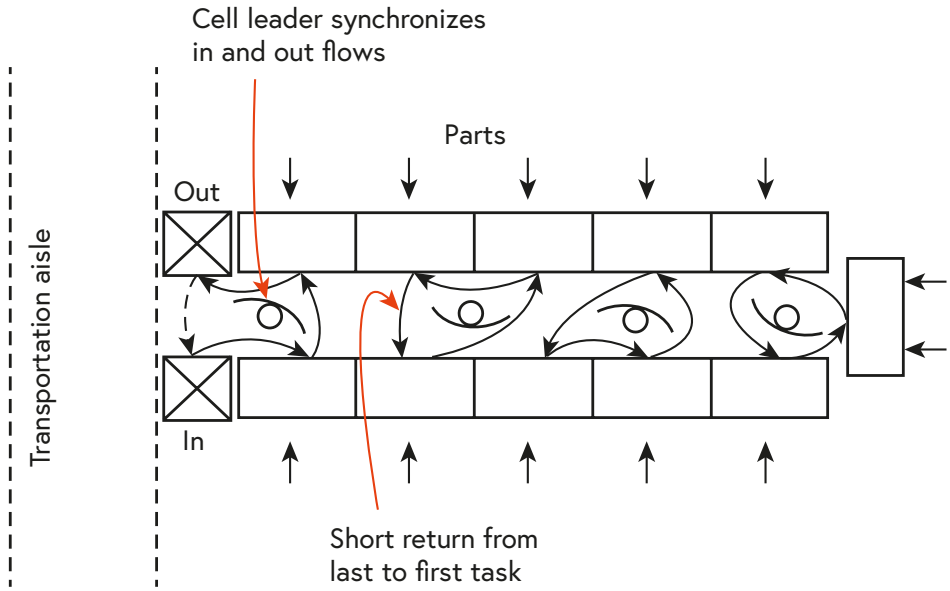


Figure 5.18 U-shaped assembly cell layout

5.4.1. The motivation for converting from a line to a cell

The spectacular performance improvements due to cells are usually based on a comparison with production in a job-shop of functional departments. While the typical improvement numbers in Table 5.1 sound too good to be true, the estimates are actually quite conservative. Managers usually need to have seen them materialize before they believe them.

Table 5.1 Typical performance improvement after converting from functional layout to cell layout

Metric	Before	After	Change	Improvement
Work-in-process (units)	500	50	-450	90%
Lead time (in work hours)	100	10	-90	90%
Units per operator per hour	5	7	+2	40%
Space occupied (square feet)	1000	600	-400	40%

Compared with production in a job-shop of functional departments, manufacturing cells have the following advantages:

- U-cells improve operator productivity and flexibility by:
  - Letting operators attend to multiple machines without walking long distances.
  - Allowing adjustments in the number of operators when the volume of work changes.
  - Allowing operators to assist each other and exchange information.
  - Encouraging teamwork.
- U-cells reduce WIP by:
  - Collocating work previously performed at different locations, thereby eliminating moves between departments and queues at entry and exit points.
  - Moving single pieces or small lots between machines within the cell.
- U-cells improve quality by:
  - Reducing lead times through the cell operations and thereby failure detection delays.
  - Making teams of operators responsible for multiple process steps and making the consequences of misprocessing visible.
- They reduce floor space requirements by increasing manufacturing density.

While local to a process segment, the improvements achieved in a cell also make nothing else worse, be it in the rest of the process or in other processes. As a consequence, they represent a net gain, not just for the process segment but for the plant as a whole.

The comparison against a functional layout is however unfair. We must be careful to separate the improvements due to the use of an assembly line – of whatever shape and length it may be – from those due to the specific cell structure.

Over other types of assembly lines, assembly cells have the advantage of being operated by a small, cohesive team collectively responsible for and with control over a much more substantial amount of assembly work. This improves not only job satisfaction and morale, but also raises quality consciousness. As already mentioned, even people who enjoy putting things together dislike assembly line work, and teamwork in a cell makes it more bearable. It is in fact so effective that it is emulated as much as

possible on longer lines by assigning segments to teams with all the trappings of a cell team except for the autonomy afforded by the cell structure.

In addition, U-shaped cells have the following advantages:

- Assembler jobs can be designed as loops through multiple stations and machines, with a short walk from the last task back to the first.
- The input and output buffers are side by side, facing a transportation aisle. This supports the following:
  - Logistics can bring new incoming units and pick up finished units on the same trip.
  - The cell leader can respond to accidental changes in flow, simply by not releasing more units into the first operation when the output buffer is full.
  - Seeing incoming and outgoing units side by side reinforces the team's sense of responsibility for quality.
- The number of assemblers working in the cell can be adjusted easily when demand rises or drops as illustrated in Figure 5.19.

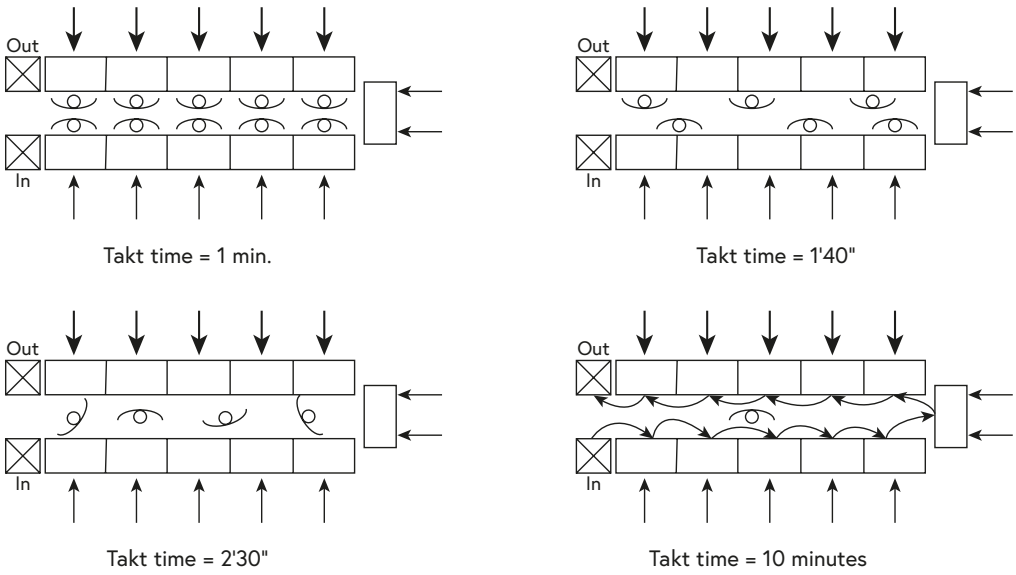
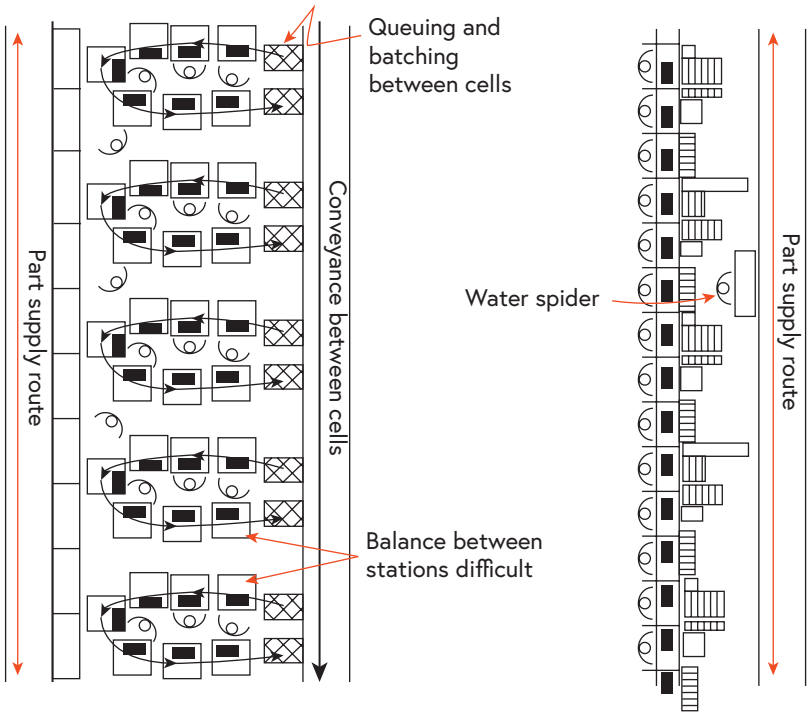


Figure 5.19 Staffing flexibility in a U-shaped cell

In principle, the same long assembly process could be implemented as a series of cells or as a single line, as shown in Figure 5.20. A series of cells is between two transportation aisles. The left-side aisle is used to deliver parts; the right-side aisle, to transfer the output of each cell to the input buffer of the next one.

Cell leaders take care of moving the parts from the transportation system interface to the assembly stations of the cells. Alternatively, we can do the same work in one single, straight line, avoiding the buffers and transportation between cells but losing the advantages of U-shaped cells and cell teams. There is only one transportation aisle, for part deliveries to flow racks and similar devices bringing parts within arm's reach of assemblers. The movement of the product itself is taken care of by the assembly line's conveyance mechanism.



a. Assembly through a sequence of cells **b. Assembly in one long line**

Figure 5.20 One long line or a sequence of cells?

5.4.2. **Limits of the U-cell**

Cells are not a panacea. There are many production operations that cannot be set up as cells. However, wherever cells are feasible, they perform so well that manufacturers should be motivated to identify where cells fit. When products grow in size in at least two dimensions, there comes a point beyond which picking across the assembly stations becomes impossible, and, in this case, a U-shaped cell cannot work, because parts cannot be delivered inside the operator work area.

Another consideration when deciding whether to use the U-shape, is the balance between assembly work and part supply. Mass market toiletries and other personal products are packaged at takt times under 1 second in automatic lines, and the bulk of the work done by people is feeding materials to automatic machines. In this case, the advantages of the U-shape are moot, and there is no point in having an operator work area in the middle of a cell when the only purpose for which it will be visited is trouble-shooting machines.

Beyond a certain product size, also, assembly work proceeds simultaneously on two sides of each unit, which is also incompatible with U-shaped cells. Cars are a case in point: At each station, there is an assembler on each side. Car engines, on the other hand, are borderline cases. In some companies, such as Toyota, Volvo, or Audi, engine assembly is single-sided, and the engines are rotated as needed for access. In other companies, such as Honda, there are operators on both sides (Figure 5.21).



Figure 5.21 Engine assembly from one side (Credit: Audi) and two sides (Credit Honda)

Stop and Think 5.3
Why are car engines not assembled in U-cells?

5.4.3. Job design in cellular manufacturing

The following three approaches are used to design operator jobs in U-shaped cells:

1. Baton-touch
2. Caravans
3. Bucket brigades

Baton-touch

Baton-touch is the most common approach used in U-shaped cells. In this mode, a job makes one operator move through a loop of tasks in the same direction as the work pieces, and in such a way that movement is either to the next station on the same side of the cell, or across to the other side. As shown in Figure 5.22, there are at most two crossings between sides. There are exceptions to this, for example machine-tools with long cycles that require operators to go back and forth to the same machine several times during one cycle.

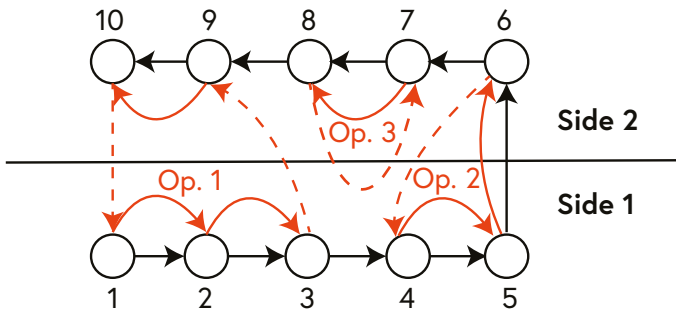


Figure 5.22 Baton-touch task assignments

Caravan

Caravan, or "rabbit-chase," is popular with operators, but limited to cells with at most two operators and loads both equally. In this case, each operator works on all the jobs in the cell in succession, picking up at each operation the last part processed through it and loading it into the next machine.

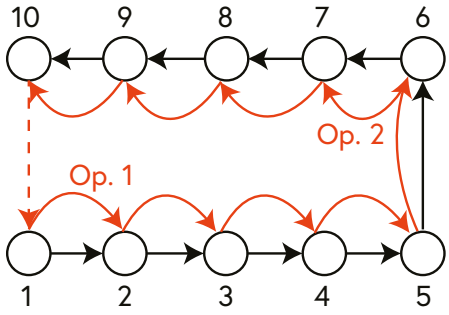


Figure 5.23 Caravan job design in cells

Bucket brigades

The bucket brigade, also known as "bump-back system," is the third approach, and the only one that does not lend itself to a work combination chart, because job assignments are not fixed. It is mostly useful in mass customization settings, such as sandwich making at Subway or custom bag assembly at Timbuk2 designs. Operators are arranged slowest to fastest along the line. Every time a unit is finished at the cell, each operator takes over the next unit from his or her predecessor and the first operator starts a new work piece, as shown in Figure 5.24. This system has been shown to be self-balancing, with all operators tending to take on work that requires the same amount of time.

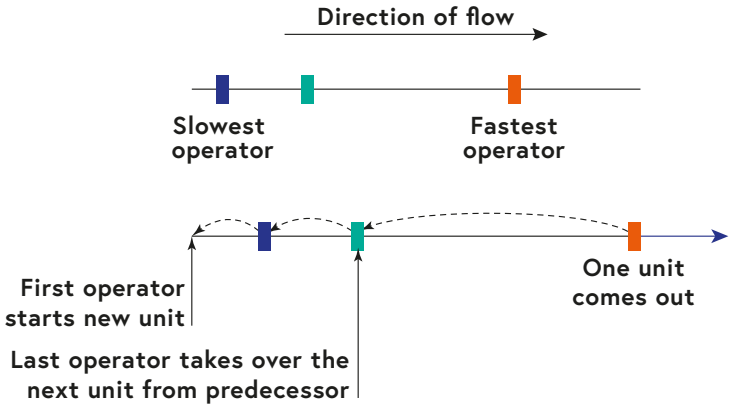


Figure 5.24 Bucket brigade operations

5.5. Assembly station design

The assembly station is the point where materials, fixtures, tools, machines, gauges, methods, and assemblers converge to make one assembly operation happen.

5.5.1. Designing assembly operations

While machining processes are *sequences* of operations, assembly processes are trees, with each branch being a sequence of operations consuming materials, feeding higher-level branches of the trunk leading to the root that is the finished product. From a part supply perspective, the assembly bill of materials (BOM) is a *graph* that is not necessarily a tree because the same part can be used in multiple subassemblies and the same subassemblies in multiple products.

An assembly process encompasses the following:

1. A subassembly structure, focused on branches and nodes and, whenever possible, shown in graphic form.
2. For each branch, a table of operations, with the following elements for each:
    a. A name
    b. Drawings, photographs, or videos
    c. Step instructions
    d. Key points
    e. Tools required
    f. Components and quantities used
    g. Labor required

Collecting this information, particularly the labor requirements, is a challenge for a production line *that does not yet exist*. The traditional method is to use predetermined times standards like *Methods Time Measurement* (MTM),[1] released in 1948, or *Maynard Operation Sequence Technique* (MOST),[2] released in 1972. The MTM process language is a worldwide standard for work design.

MTM and MOST allow you to assign times to sufficiently small increments of manual work based on tables, and add them up. Doing this for work that has not been done yet, however, is both difficult and tedious, and when actual production starts, the results are commonly within ±30% of the estimate.

One solution to establish the need for assembly time is to, whenever possible, take apart and reassemble a unit of product several times. At Honda, this practice is called Zenbara. Figure 5.25 shows an example of a four-cylinder engine. Of course, it is easier to do Zenbara with a fishing reel than with a port crane or a jumbo jet. A Zenbara exercise helps in defining a subassembly structure, identifying tooling requirements, sequencing the assembly work, and estimating how long each step takes. Several different options should be explored.

1 Maynard, H. B., Stegemerten, G. J., & Schwab, J. L. (1948). *Methods-time measurement*. McGraw-Hill.

2 Zandin, K. B. (2003). *MOST work measurement systems*. Marcel Dekker.

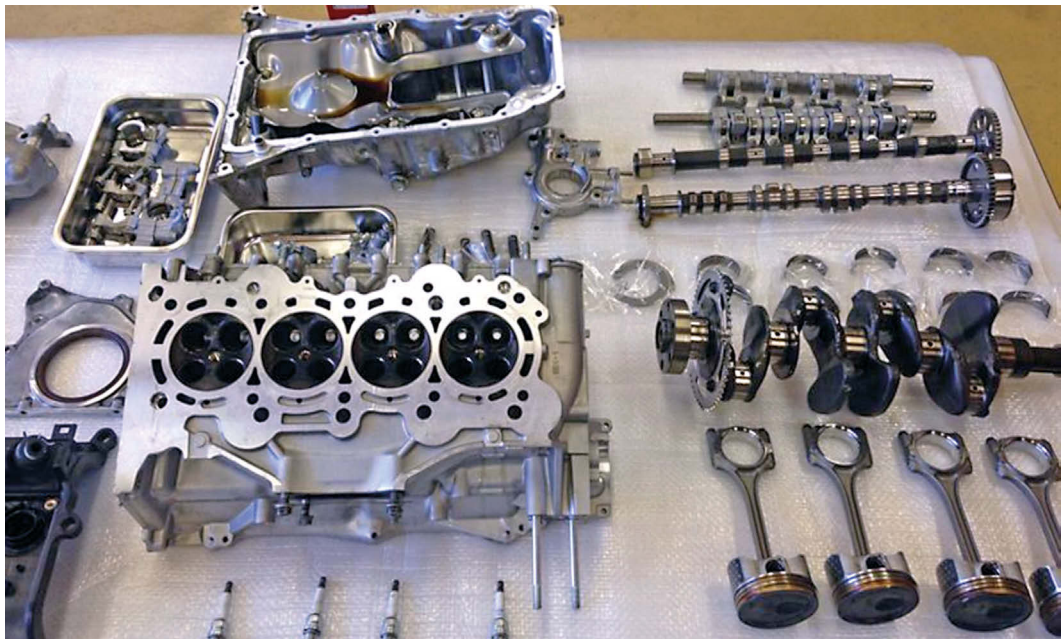


Figure 5.25 Zenbara informs assembly process design at Honda

The next step in process design translates the abstract assembly tree into a network of workstations with a conveyance mechanism between consecutive stations, tools, fixtures, part presentation or feeding devices, and work instructions. The flow of materials of these designs is evaluated based on blueprints where toy-scale versions of the products are moved, as in a board game. This is sometimes done with software discrete-event simulations, but the cost, particularly the engineering time spent developing and tweaking simulation models, is often not justified.

These tools, however, do not simulate the human experience of working on the line. This can be done with full-size cardboard mock-ups or using Virtual Reality (VR) or Augmented Reality (AR). These physical or digital mock-ups allow operators to provide feedback and improve the design, as well as the opportunity to conduct more realistic time-and-motion studies.

Before final implementation of the assembly concept, a trial line or cell can be set up in a dedicated area where final adjustments are made before the line or cell is built for production ramp-up at the factory floor. The trial line can be used for training and preparing operators.

5.5.2. **Assembly logistics**

The availability of materials in the right quantities, at the right assembly stations, at the right time, and presented for assembly convenience is a central concern in assembly but not in machining, where you have one workpiece flowing through an entire process and the most cumbersome logistical issue is chip disposal. The number of parts ranges from dozens for a fishing reel to millions for commercial aircraft, all of which must be made or bought, transported, stored and retrieved, and delivered to production.

The designers must then plan for every part, from the assembly station back to the suppliers, whether internal or external, answering questions like:

- Should the parts be kitted or presented in racks behind the assembly station?

- How often should they be delivered to assembly?
- How often should they be delivered to the plant?
- Is a buffer needed to decouple internal and external lead times?
- Are suppliers available within 1 hour drive of the plant?
- How do you organize ordering and deliveries for local and remote suppliers?

As part of designing the process, the manufacturer must decide which parts to make in house and which to buy. In the business literature, this choice is supposed to be based on whether the part embeds the company's "core technology." In practice, this criterion is little help, because deciding what is core technology is no easier than deciding what to make or buy. Usually, the motivation for making parts in-house is to control the supply and for outsourcing to reduce costs.

Companies bring the production of some items in house by acquiring suppliers, and increasing costs by extending their more generous benefits to the supplier's employees. The benefit of this acquisition, however, was to deny competitors access to this supplier's capabilities and production capacity. In boom times some years later, the same company may partially outsource production of some of the former supplier's items, in order to acquire surge capacity without investing in more machines.

Buying an item is only possible if there are suppliers. There may not be any if the item is based on state-of-the-art technology. There may not be any local suppliers if the assembly plant is located in an area that has no manufacturing tradition. Designing around unavailable materials is a professional hazard for product and process designers. An old engineering joke is to call such materials unobtainium. In James Cameron's science fiction movie Avatar (2009), humans tried to colonize the distant planet Pandora specifically to mine the fictional unobtainium.

There are many details about supply chains and logistics that is covered in later chapters. The key point is that it cannot be ignored when designing an assembly process.

Stop and Think 5.4
Why is logistics one of the defining characteristics of assembly processes?

5.5.3. Assembly stations for small products

"Small," in this context, means small enough for the assembler to pick parts across the station. Products in this category include most manufactured goods for consumers – such as electrical razors, hand tools, toys cell phones, and cameras – as well as subassemblies or spare parts for larger products – ranging from shock absorbers for cars to cockpit switches for airliners. The small product size creates opportunities in station design that are not available for larger products, such as the following:

- Since work takes place on one side only, such stations can easily be arranged into U-shaped cells with the assembler work area inside.
- Part deliveries do not interfere with the work, and stations can be arranged to face transportation aisles.
- Many small products are light enough to be handled manually.

Figure 5.26 shows a modern assembly line design at a Hilti factory in Thüringen, Austria. In these work stations, the engine for an industrial saw tool is assembled. Note the projections on the operator's hand

and the work table that show digital work instructions and assembly steps using AR projection systems from the Belgian company Arkite. Also, note the yellow screwdrivers mounted to automatic recoil systems available right above the operator's head and the assembly material that is supplied from the transportation aisle behind the station. The screwdrivers measure torque automatically and count the number of screws mounted.



Figure 5.26 Assembly stations for hand tools (Credit: Hilti)

Most manufacturers have potential to exploit more of these opportunities. The most common concern is station size. You do not need a 60"×30" bench to work on a product whose longest dimension is 6 inches, and using these unnecessarily large stations causes the following problems:

- Neighboring assemblers are too far apart. Their communications are reduced, they are unable to relieve each other, and one assembler cannot conveniently work at more than one station when volume drops.
- The excessive distance between stations complicates conveyance and hinders the product flow.
- Just because it is there, the extra bench space fills up with unnecessary WIP, tools, fixtures, and other items unrelated to the operation at hand.
- The floor space used by a line of these large stations is twice or three times more than it should be.

There is also often a good potential in better use of conveyors. Conveyors often move parts that are small and light enough to be passed around by hand, for the sole purpose of enforcing pace. The assembly station is then frequently located *off* the conveyor, that is, not integrated as part of the conveyor. With short operation cycles, the additional handling can easily account for 40% of the assemblers' time. Such benches are used today, not because they meet the needs of assembly, but because they are available off the shelf from multiple suppliers.

There is a product size beyond which picking across an assembly station is no longer an option. This limit is not expressible as an exact number, but it is easy to see, for example, that cell phone assembly is on one side of it and car engine assembly on the other. With large products, it is also possible and frequently advisable to have assembly operations proceed on both sides of the conveyor.

Box 5.3 **Apple's iPhone assembly by Foxconn and Wistron**

Assembly of small electronic products are today mostly done by *Electronics Manufacturing Services* (EMS) firms. Since the 1970s, EMS firms design, manufacture, test, distribute, and repair electronic components and assemblies for Original Equipment Manufacturers (OEMs). When producing the designs of other firms, the correct term is *contract manufacturing*. Some of the most known EMS/CM are Foxconn, Jabil, Flextronics, Solestron, Pegatron, and Wistron. Of these, Foxconn, Pegatron and Wistron feature in media for producing Apple's iconic iPhone.

Seeking lower production costs than they currently find in China, contract manufacturers have recently looked to India for expansion. Wistron, for example, has made iPhones for Apple in India since 2017. In August 2021, it opened an iPhone factory in Kolar, outside Bangalore.[3] Similar as for Foxconn's facilities in China, the media coverage is not all positive.



Figure 5.27 Smartphone assembly in Shenzen, China (Credit: Juvertson/Flickr)

These factories employ thousands of workers. For example, Foxconn reportedly[3] employs 350,000 people in both "Foxconn City" in Shenzen and Zhengzhou, Henean province, China. The same source says the Zhengzhou factory produces 350 iPhones per minute on 94 iPhone production lines, and up to half a million a day. Other sources suggest around 160,000 workers in this factory are working on Apple products. In such "factory cities," workers live and work.

In the Wistron factory in India, about 9,000 contract workers had been employed since the opening in August 2020 until violent worker protests started in December 2020 and made international headlines. Workers protested against exploitation and sweatshop-like conditions, with lower pay than promised. Involved contract workers were fined and fired, and new employees were hired. The Wistron headquarters promised improvements.

One report[4] tells the story of a worker who is doing visual quality control of the iPhones, looking for errors and scratches. "At the start, I had 36 seconds to find scratches or defects. Than 30 seconds – if it has a defect I didn't find, I would be scolded. They could track every iPhone back to me." This person worked 12-hours shifts with two 30 minutes' breaks. He inspected 1,320 iPhones a day. His wife who also worked in the factory assembled four screws on the same cycle time. Hence, a screw every 7.5 seconds until 5,280 screws were assembled on her shift. Both these job task can relatively simply be automated, and would have been if the factories were located in high-cost countries. Automated vision control with machine learning and a small assembly robot would solve these two tasks, but for the current 120–170 USD monthly salary for manual workforce, the work remains done by human workers.

3 Babst, A. & Sander, M. (2021, March 12). 1320 iPhones pro Schicht. *Neue Zürcher Zeitung*, p. 23.

4 Half of iPhones manufactured in central China's Zhengzhou city (2017, September 19). *China Daily*.

5.5.4. Assembly stations for medium-sized products

With large products, presenting a particular assembly spot to the assembler is a challenge. Sometimes, assembly work must even be carried out underneath the product. If possible, this should be avoided by providing fixtures that rotate the product to present the target assembly areas to the assemblers.

Figure 5.28 shows a good example from diesel engine assembly in a Swedish Volvo factory. The engine is hanging on a flexible conveyor that together with a floor-mounted rising fixture can tilt the engine and adjust the working height according to each operator. Note also the ergonomic floor, personal safety equipment, good lightning conditions, and tidy workplace.

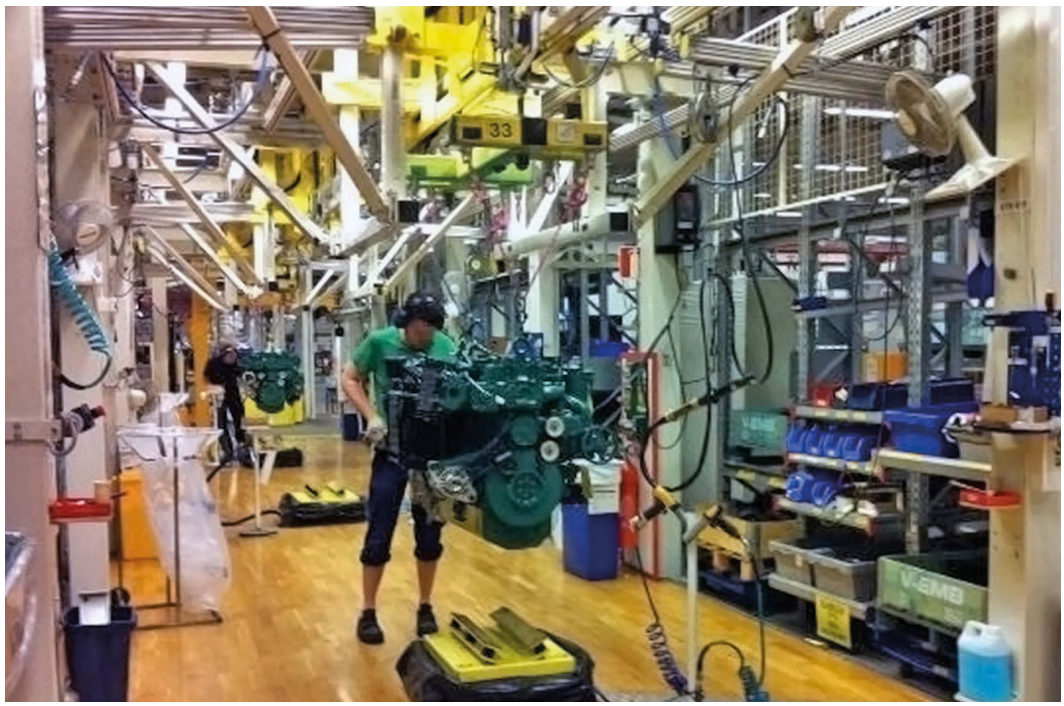


Figure 5.28 Diesel engine assembly on hanging conveyor lines in a Volvo factory in Sweden

When the product is larger, also the parts that go into it tend to be larger. This means parts should only be stored at the work station for one or a few products. If this policy is not followed, the work station will be crammed with parts, boxes, and pallets. This, however, is not unusual. Factories that have not undergone a lean transformation tend to store parts locally. As all humans, assemblers intuitively like to have a little extra material "just in case." In work designs, where assemblers are responsible for getting parts themselves, this situation is often extreme. The result is low productivity due to the need to search for parts among many, bad lighting conditions and visibility, and obsolete parts stored all over the factory.

The better solution is to make parts flow to the place of assembly just before they are needed. The right side of Figure 5.28 shows a good example where parts are supplied just-in-time to the work station using kanban for larger subassemblies and parts and two-bin systems for small standard parts and consumables.

5.5.5. **Stations with required dwell times**

Many assembly processes include operations where parts must remain in one place, to cool off, until glue is dry, to let paint dry, or until sealant is cured, for a period that exceeds the takt time. The challenge then is to design stations through which the units flow, but where they come out right next to where they go in, so as to allow one operator to handle both loading and unloading while possibly taking care of upstream and downstream stations as well.

As shown in Figure 5.29, tunnel ovens support first-in-first-out but they keep assembly stations apart. A shuttle oven, on the other hand, keeps the stations close together but is a last-in-first-out device. Hence, when the product characteristics allow, the tunnel oven is preferred.

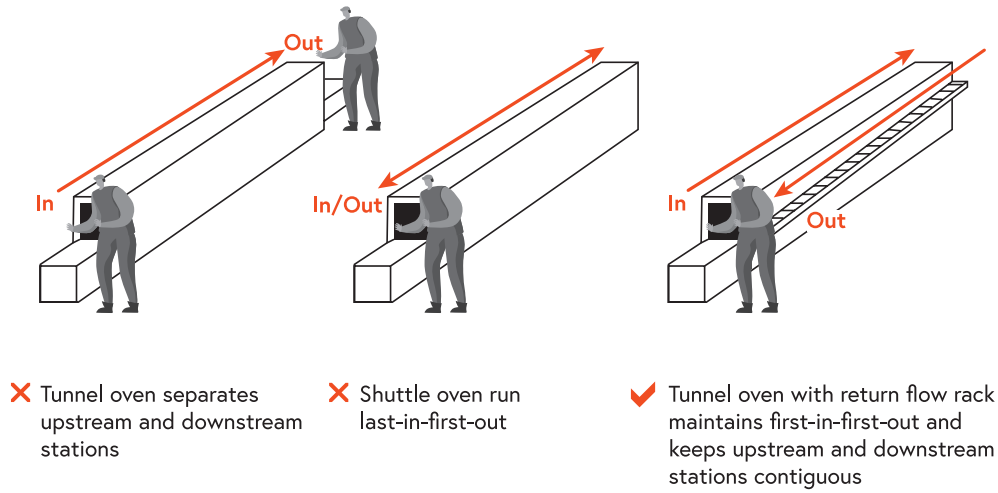


Figure 5.29 Three different curing station designs

5.5.6. **Location of fixtures**

The ideal assembly fixture lets the assembler work with both hands, allows all assembly steps to be done in an ergonomically appropriate way, docks as needed in automatic equipment, is also used as a kit pallet for product-specific parts, and, if possible, can hold more than one product. Fixtures for manual assembly travel through multiple stations, and need to return afterwards to the first station, traveling above or below the assembly line.

Fixtures used for pressing together or welding parts do not travel and hold the parts in place with clamps or pneumatic cylinders. The station design issues with these type of fixtures are to make manual loading and unloading easy, fast, and safe. Fixtures must be oriented for easy loading and unloading. The machine must be prevented from starting while the operator's hands are inside, but the most common approach – dual finger switches – is a productivity drag in short cycle jobs, and equally safe but more productive alternatives need to be sought on a case-by-case basis. In-line mechanical automation is an alternative to off-line robotics that is cheaper in high volume applications and relies on jigs and locating pins to align components.

Manual assembly fixtures differ from machining fixtures in several ways:

- They are not subjected to the forces of cutting tools.
- Whereas machining fixtures mostly circulate between a rack and one machine, manual assembly fixtures often carry a unit of product through multiple stations.
- They must be "people-friendly" and must not interfere with assembler movements and in particular be free of sharp edges and trip hazards.

The most effective fixture is not always the most complex one. In low-volume, high-mix computer assembly, for example, the simple *Lazy Susan fixture* of Figure 5.30 was all that was needed. For all models, the assembly process consists in populating a chassis from the top with such parts as power supplies, disk stacks, printed circuit boards, plug-in components and connecting them with cables. The Lazy Susan allows the assemblers to always have the point of assembly within easy reach.

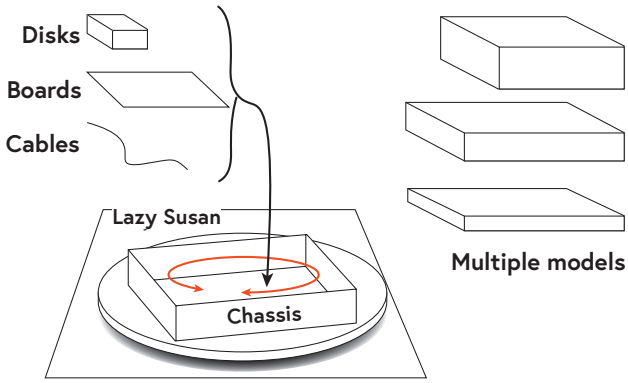


Figure 5.30 Lazy Susan for flexible computer assembly

It is often not possible for a single fixture to serve from the beginning to the end of the line, and product units have to be transferred between fixtures two of three times along the way. For each fixture used through multiple stations, a means must be provided to return from the last to the first station where it is used. In a U-shaped assembly cell, light fixtures can be hand-carried from the last station to the first, and heavier fixtures moved on a cart. This does not work, however, at short takt times in the order of 15 to 30 seconds, in which the transfer must be automated.

5.5.7. Location of hand tools

Common hand tools used in assembly include the following:

- Pneumatic or electric screwdrivers and nut runners
- Wrenches and torque wrenches
- Fluid dispensers: Cleanser, adhesive, sealant or flux
- Soldering irons
- Pliers and crimping tools
- Q-tips and brushes
- Mallets and hammers

As explained in Box 5.4, hand tools should be attached to the station and not to the assembler.

Box 5.4 **Should operators have personal tools?**

The first priority with hand tools is that they should be attached to the station and not to the assembler. The main reason for this is the need for process consistency. If each assembler uses private tools, these is no way to assure that the job will be done the same way every time. In addition, hand tools attached to the station can be tethered to power sources and electronic controls.

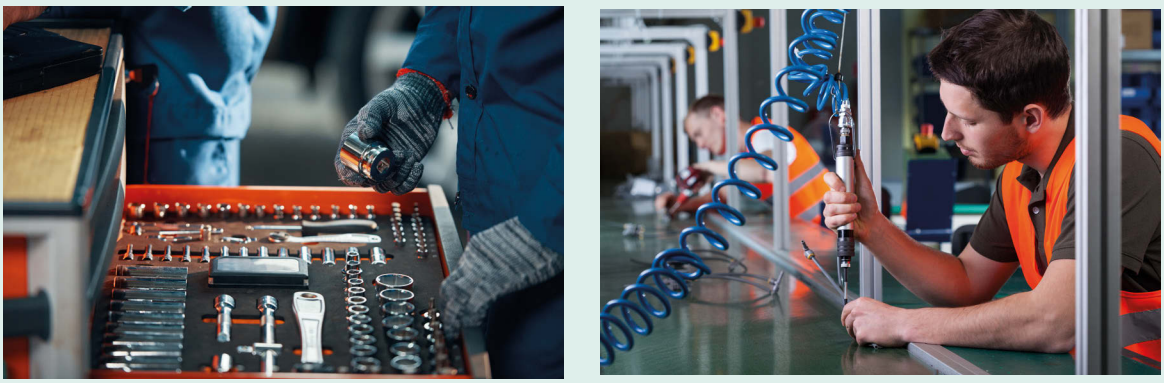


Figure 5.31 Tool box on wheels (left) and assembly station with mounted tools (right)

Private tool boxes in sizes ranging from shoe boxes to wardrobes on casters are a common sight in assembly shops, primarily in the aerospace industry. On some car assembly lines, assemblers wear tool belts, which are less bulky than tool boxes, provide easier access to tools, and are much less likely to become repositories of personal items unrelated to the work. From the point of view of process consistency, however, tool belts are no better than tool boxes, since they also carry private tools.

The use of private tool boxes reflects a confusion between the needs of maintenance technicians and production operators. Maintenance technicians gear up for contingencies. When technicians open a machine, even for routine maintenance, they need to be ready for whatever they may find.

In addition, the tools they use are not needed at each machine on a routine basis. It therefore makes sense for them to carry a toolbox around. Production operators, on the other hand, execute the same routine with the same few tools over and over again at the same locations. Extraneous tools are not wanted, because they may interfere or be confused with the right ones. Assemblers should be able to approach an assembly station with nothing more than clothing and safety gear on their person, and find all the tools they need at the station.

The location of hand tools is an area where the assemblers themselves have the most expertise. They should always be consulted, and, whenever possible, involved in making decisions about it. While assemblers are generally not that concerned about such issues as plant-wide material flows, tool location affects their immediate daily work experience.

Locating a tool means positioning it within arm's reach of the assembler – with the most frequently used tools the closest – and orienting it in such a way that the assembler does not have to turn the tool 180° or even 90° prior to using it. Screwdrivers used for vertical screws are pencil-shaped and hang vertically on balancers; those meant for horizontal screws are gun shaped and hang horizontally on balancers. Figure 5.32 shows an example of tool placement on a station for small product assembly.

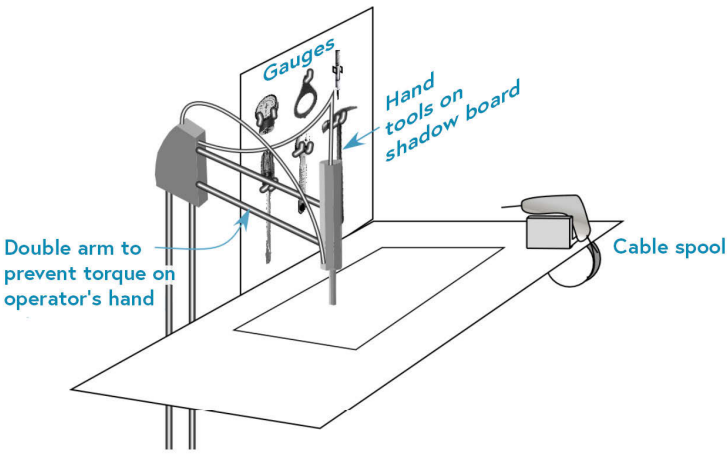


Figure 5.32 Tool placement at a workstation

For large products, particularly on moving conveyors, the challenge of attaching tools to stations is met by mounting the tools on moving or "nagara" fixtures that follow the product through the station and automatically return to the starting position afterwards. It is also known as "flying carpet" as in Figure 5.33.

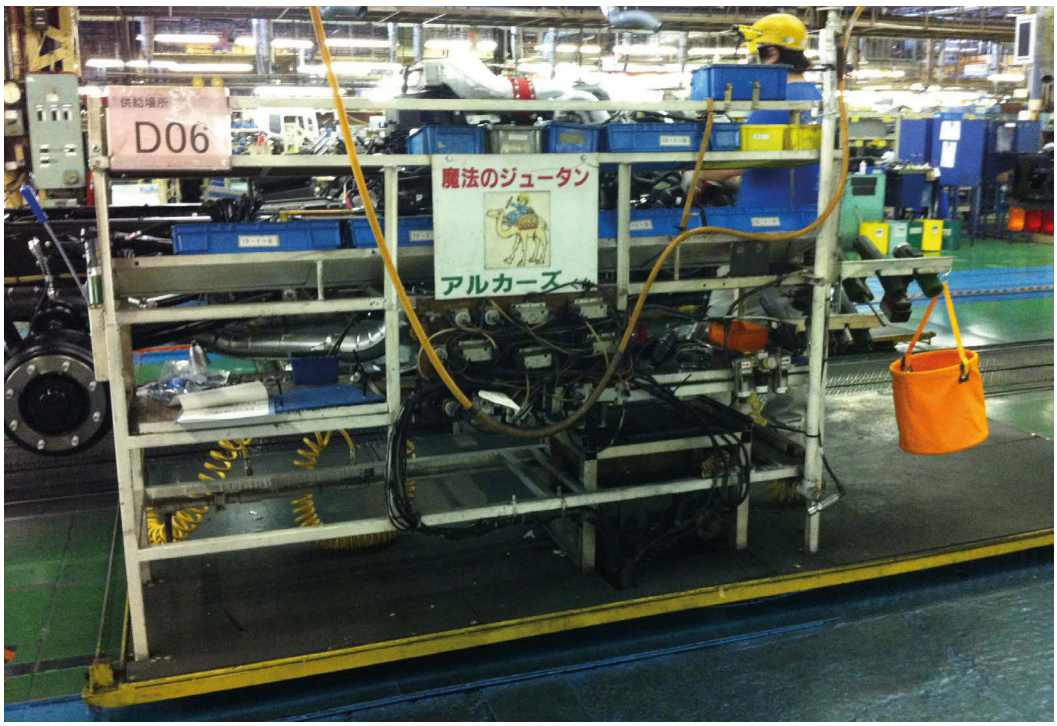


Figure 5.33 Nagara "flying carpet" fixture in an automobile assembly factory in Japan

5.5.8. **Automated assembly and packaging**

When takt times fall below 15 seconds, humans are not reliable resources for assembly. When takt times fall below 5 seconds, humans are incapable of following the pace over an extended period. In these cases, automation is called for. The same is true when the tasks are too heavy or difficult for humans to get right.

> **Definition**:
> Automated assembly refers to the use of mechanized and automated devices to perform assembly tasks in an assembly line or cell.

Automated systems are easiest for tasks that are repetitive with little variation. Therefore, the vast majority of automated assembly and packaging are tasks with limited variation. As more intelligent automation appears, tasks with a higher degree of variation are also suitable for automation.

As argued, all repetitive assembly tasks are ultimately suitable for automations. Yet, the solution has to be economically feasible for manufacturers and have a reasonable pay-back time; three years is often seen as good, five years is ok, and anything longer requires elaborate analysis. Figure 5.34 shows an example where a collaborative robot is working with a human in a BMW axle drive assembly factory.



Figure 5.34 Axle drive assembly with collaborative robot in BMW's Dingolfing factory, Germany (Credit: BMW Group)

Automation has many benefits over human labor. If programmed well, robots do not make mistakes, and they can work 24/7 without rest and toilet breaks. They can work considerably faster than humans, can adjust their pace and rhythm fast, and do not complain. They are also stronger than humans. Modern robots need less floorspace than humans. They also automatically track and store data.

Yet, automation is not a panacea. In any case, there will be humans in the loop, who program, control, and troubleshoot the robot. These jobs have to be done by trained engineers, who are more expensive than assembly operators. Robots are also not (yet) good at taking their own decisions informed by the context they work in. Most strikingly, robots are useless in improvement work.

5.6. **Disassembly**

Disassembly is the opposite of assembly. While it has received much less attention than assembly in the manufacturing literature, it is the main process type for some companies. Examples are slaughterhouses, yards for decommissioning of infrastructure such as nuclear plants, oil platforms and ships, and recycling plants that extract valuable parts from products that have reached end-of-life. In a strict sense, none of these processes are covered by the definition of manufacturing set forward in Chapter 1, but the point is that, often, these processes are part of a larger production system that make use of the parts for subsequent assembly or mixing processes.



Figure 5.35 Examples of disassembly processes: Slaughterhouse, fish cutting, and watch remanufacturing

Disassembly is a distinct process type with the following characteristics and process design challenges:

- Output is defined by the input, not market demand.
- Quality of incoming materials is more difficult to control.
- Information is revealed during the disassembly process.
- Processes are more challenging to standardize and, hence, automate.

One of the defining characteristics of disassembly processes is that the output mix is determined by the input and not by the demand. The quantity of sirloin, flank steak, or brisket is determined by the cattle's anatomy, not the public's appetite for the various cuts. We can say that disassembly processes have a "reversed bill of materials," where the exact amounts of the resulting parts are not defined. Note that this makes disassembly processes follow a push design: When the disassembly process is started it usually continues until the different parts are separated. If the customer orders a tenderloin, there is no way to extract only that part of the cattle without starting the whole disassembly process.

The next defining characteristics are that it is more difficult to have high quality control of incoming material and that information about the state of the material is revealed during the disassembly process. For example, the exact anatomy of a cattle is first learned when the animal is cut open. Some cattle get more stressed by the transportation and slaughtering process than others, which affects the quality of the meat. There is also natural variation in meat quality depending on both heredity and environment. The same is true for products that are used and come back for maintenance, repair, overhaul, or remanufacturing. One does not know the state of the inner parts of a used product before opening it.

For the above reasons, it often difficult or impossible to automate disassembly processes. For example, the meat cutting processes remains largely manual work because it requires the skills of a trained slaughterer to consider each cattle's anatomy. Exceptions to this is disassembly of poultry and fish, as tightly controlled breeding is able to create standardized animals.

Box 5.5 **Beaching: Disassembly of ships**

When ships, oil platforms, nuclear plants, military installations, aircraft, and other infrastructure reach end of life, they are decommissioned – taken out of service. The issue is then what to do with them. The market prospects for the recycling business has not been anything close to the manufacturing of new products, so it is often treated as an afterthought for manufacturers. This "linear economy" thinking is disastrous for the environment and therefore subject to governmental and international regulation. The goal should be to reach a "circular economy," which would give disassembly a more prominent role.

Even with international regulations, there are countries that, for economic reasons, offer substandard disassembly opportunities, and companies that, for economic reasons, are willing to take advantage of it. One such practice is "beaching" of ships. Figure 5.36 shows a disassembly operation of ships on the Gadani ship-breaking yard in Pakistan. It is the world's third largest ship breaking yard with a capacity of disassembling 125 ships a year. The "yard" is a 10 km long beachfront and employs 6,000 people.



Figure 5.36 Ship breaking on the Gadani ship-breaking yard in Pakistan (Credit: Tomàs Halda)

In "beaching," each ship is gunned towards the coast and beached, providing the ship-breakers access to the prow. As the ship breaking proceeds, successive tides move the wreck forward, providing further access. The beach then is lined with shops that organize the resale of steel plates, furniture, and appliances recovered from the ships. The work of breaking the ships is done manually and is both dirty and unsafe, as the actual content of the ships is generally unknown. The workers' salary is low.

5.7. **Remanufacturing**

> **Definition**:
> Remanufacturing is the industrial process of returning a used product to the function and appearance of the original product.

Remanufacturing is an industrial process that starts with the used product and creates a new product that is similar to, or better than, the original pre-use product. Remanufacturing is therefore a special case that combines disassembly and assembly processes. In many ways, this combination makes remanufacturing very complex to manage effectively and efficiently. As shown in Figure 5.37, remanufacturing is distinctly different from recycling, reuse, and repair activities. Refurbishing is another term for remanufacturing, which relaxes the requirement to bring the product back to like-new quality.

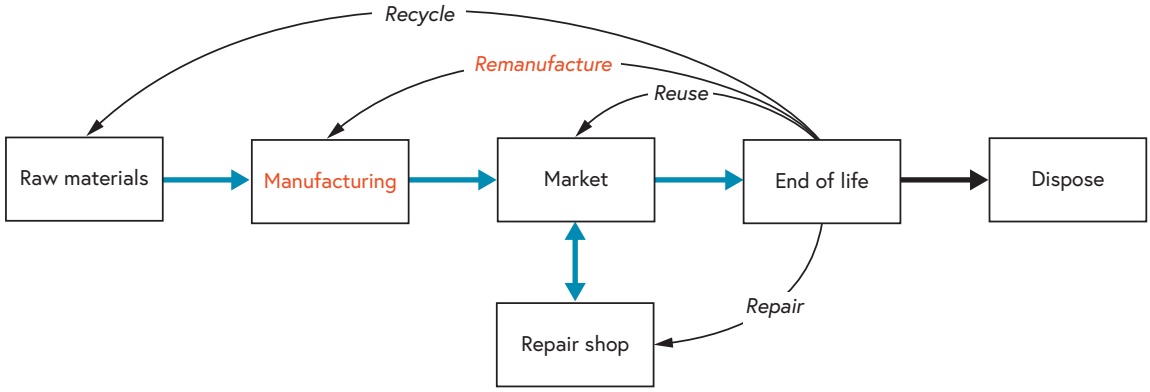


Figure 5.37 Remanufacturing returns a used product to the specifications of the original product

Remanufacturing is as old as manufacturing. While many manufactured products have the potential to be remanufactured, few are. The European Remanufacturing Council suggests that the ratio of remanufacturing to new manufacturing of products that are suitable and potentially economical to remanufacture or repair is 1.9%.[5] There is a great potential for remanufacturing as a manufacturing growth sector, but the development is slow. Still, it is by no way not negligible; the same report suggests that 190,000 people worked in remanufacturing in Europe in 2015, generating a revenue of €30 billion.

As indicated by the low ratio of remanufacturing to manufacturing the economic incentives to remanufacture is often absent or low. Both revenues and costs may be less attractive for manufacturers:

- On the revenue side, customers are generally not willing to pay the same or more for a "new" old product than for a new product. They usually require a large discount for remanufactured products. Often, customers judge the quality of remanufactured products as lesser than for new ones – even though the opposite may be true. For this reason, there can be fewer potential customers for remanufactured products.
- On the cost side, the remanufacturing processes are very hard to plan and control efficiently. Due to greater product variation, a skilled workforce is needed and automation is often impossible. Remanufacturing usually has a much lower economy of scale than manufacturing operations. In many markets, additional certifications add extra administrative costs for remanufactured products. An example of this is medical equipment. The only clear cost benefits are (1) lower cost for

5 Parker, D., Riley, K., et al. (2015). *Remanufacturing market study*. European Remanufacturing Network. EU report.

the "core," that is, the part that is remanufactured, and (2) fewer assembly steps may be needed if subparts are in excellent conditions.

For some products, however, there is a positive return on remanufacturing. The following products are remanufactured mostly because *the core is more expensive to manufacture than remanufacture*:

- Engines, transmissions, compressors, and other components in automotive, construction equipment, rolling stock, marine, and aerospace
- Military equipment
- High-end or vintage furniture
- Medical equipment
- Electrical and electronic equipment, computers, and phones
- Batteries
- Printing equipment and toner cartridges
- White goods and restaurant and food-service equipment
- Manufacturing machines
- Musical instruments

How remanufacturing activities are organized is another question. Usually they are conducted in separate facilities, either owned by the OEMs or owned by third-party contract remanufactures or independent operators. In general, high-end products that require specialized proprietary knowledge are organized in-house by OEMs, while more generic products are organized by external companies.

Stop and Think 5.5
What are the operational factors that makes remanufacturing a greater challenge than original manufacturing in assembly factories?

Consider the example of truck engines (Figure 5.38). The cost of a remanufactured truck engine is approximately 30–60% cheaper than building a new one. When we know that every engine block has the potential to be reused three times, we should expect the produced volumes of remanufactured engines are approximately three times as high as for new engines. Unfortunately this is extremely far from the current reality. Why? Because remanufacturing engines is a cost- and resource-intensive activity with considerable risks related to effectiveness and efficiency. The unknown variation in incoming product quality and states is making it very difficult to design an efficient manufacturing process. Future integration of more sensors in products and real-time product-state monitoring may help shorten lead times and reduce the complexity of disassembly and product diagnosis.

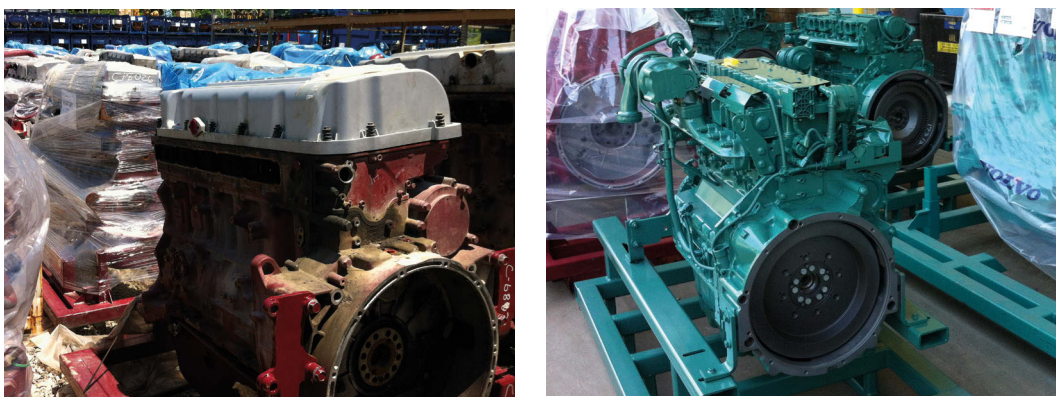


Figure 5.38 End-of-life truck engine versus remanufactured engine

Remanufacturing requires highly knowledgeable employees, who are well trained in the technical elements of the different products remanufactured. Knowing what to discard and what to keep is a skill not easily gained. Figure 5.39 show a high-level overview of the typical process steps in remanufacturing. In reality, the process flows are much more complex, with back flows and loops.

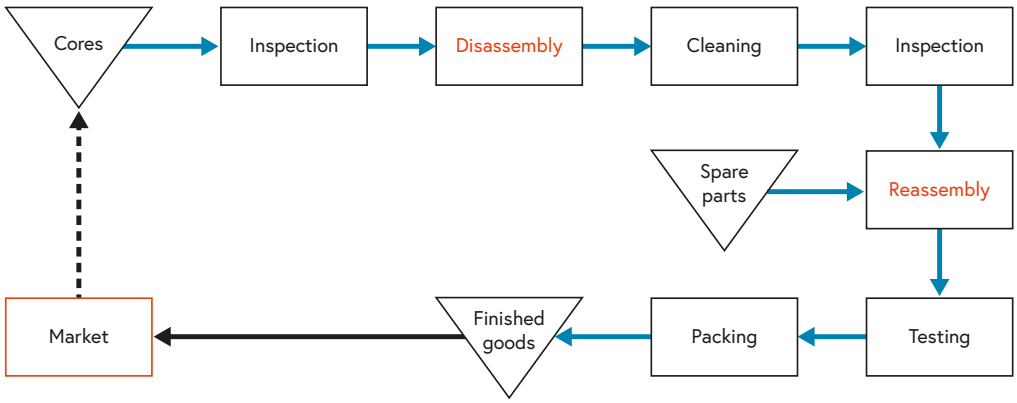


Figure 5.39 Typical processes in a remanufacturing operation

Due to the complexity involved in judging quality and the technical expertise needed, the average age of the employee stock in a remanufacturing factory is often 20 years higher than in the factory assembling the new products on a production line (see Figure 5.40, left photo). Their total knowledge gives much flexibility, but on the expense of efficiency. In remanufacturing, it is often difficult to split tasks on operators and a grand challenge to balance a production line. Flexibility, however, is a necessity in this type of production that is particularly characterized by high variability in several ways:

- You don't know your actual raw material stock because it is hard to judge its quality: You first now the condition of the incoming materials after disassembly and washing.
- You need much judgment to decide if you should remanufacture or scrap a part: It is hard to codify all needed knowledge.
- You don't know your process times: For engines, it can typically vary from 4 minutes to 4 hours.
- You often remanufacture all models from all times: Products from the last 30–50 years may be included, making the variants handled grow year-by-year (see Figure 5.40, right photo).
- There is built-in waste in the process: You necessarily process many parts that will be scrapped at a later stage when flaws become apparent.



Figure 5.40 A remanufacturing facility for truck engines and transmissions in the US

5.8. **Line balancing with Yamazumi charts**

The most common graphic tool for assembly line balancing is the Yamazumi chart, as shown in Figure 5.41. "Yamazumi" (山積み) is not a person's name. It means "pile stacking." In a Yamazumi chart, each task assigned to an individual station has a bar with a fixed width and a height proportional to its duration. You stack these bars into a column for the station.



Figure 5.41 Balancing operator jobs to fit below the red line in a Yamazumi chart

The red line materializes the takt time. If the pile crosses it, the operator's job at this station is impossible. Either the tasks must be reengineered for shorter durations or tasks must be reassigned backwards or forwards on the line. Conversely, if the pile is too short, it means that an operator will have slack time.

In the Yamazumi chart, you try to fill up the takt time at each station, except the ones assigned to *team leaders*, so that they are able to attend to the leader tasks. The team leader stations are also the primary target for improvement, as they offer the best opportunity to eliminate a station altogether.

When balancing assembly, many line designer try to build in small amounts of slack – on the order of 2% of the takt time – on the front-end of the process, to recover from small disruptions without hurting the line's ability to deliver one unit every takt interval at the back end.

The Yamazumi chart works with manual assembly because it is almost infinitely divisible: The work can be split into tiny increments that *can* be reassigned between stations while, with machines, the work of a lathe cannot be reassigned to a grinder. Where machines are involved, you need the work combination charts described in Chapter 4, Section 4.7.

Summary

The most essentials points covered in this chapter were:

1. Assembly is the final step in discrete manufacturing and usually the most human-intensive process in manufacturing.
2. Assembly processes can be separated into five generic types:
   - Joining, which is the central value-adding activity
   - Handling
   - Controlling
   - Adjusting
   - Special activities
3. Assembly lines need to be carefully designed in order to enable smooth flows of people, products, materials, and information.
4. Assembly cells are U-shaped assembly lines, where a maximum of 8–10 operators work inside the U and parts are fed from the outside.
5. Three different job designs in cellular assembly cells are baton touch, caravan, and bucket brigades.
6. Logistics and work place organization is key to high efficiency in assembly processes, but vary with the characteristics of the products such as volume, variety, and size.
7. Automation is an objective for many assembly processes and should be a requirement when takt times fall below 15–30 seconds.
8. Part presentation helps operators work efficiently and ergonomically.
9. Disassembly processes are the opposite of assembly processes and found in industries such as meat processing, remanufacturing, recycling, and breaking of infrastructure and vehicles.
10. Remanufacturing is the industrial process of returning a used product to the specifications of the original product. It combines disassembly and assembly processes.

Test yourself

1. What is bench assembly and how does it differ from an assembly line?
2. What characterizes human work at assembly lines?
3. What are the most common joining processes in assembly?
4. What are the two main factors in assembly performance?
5. How should parts be presented to an assembly station?
6. What are the advantages of U-cells compared to I-shaped assembly lines?
7. How do the cell job designs baton touch, caravan, and bucket brigades differ?

8 Should operators working in an assembly cell carry their own tools?
9 What are the defining characteristics of disassembly processes?
10 Why is it difficult to organize remanufacturing in an assembly line?

Further reading

Baudin, M. (2002). *Lean assembly: The nuts and bolts of making assembly operations flow*. Productivity Press.

Bebersdorf, P. & Huchzermeier, A. (2022). *Variable takt principle: Mastering variance with limitless product individualization*. Springer.

Berggren, C. (1993). *The Volvo experience: Alternatives to lean production in the Swedish auto industry*. Macmillan.

Hyer, N. & Wemmerlov, U. (2001). *Reorganizing the factory: Competing through cellular manufacturing*. CRC Press.

Lotter, B. (2013). *Manufacturing assembly handbook*. Butterworth-Heinemann.

Nasr, N. (Ed.). (2019). *Remanufacturing in the circular economy: Operations, engineering and logistics*. John Wiley & Sons.

Nof, S. Y., Wilhelm, W. E., & Warnecke, H. J. (2012). *Industrial assembly*. Springer Science & Business Media.

6 Automation

Former U.S. President John F. Kennedy once said "Automation does not need to be our enemy. I think machines can make life easier for men, if men do not let the machines dominate them." In Kennedy's day, in this context, "men" meant "humans." He addressed the common but as yet never realized fear that automation leads to massive unemployment or to the enslavement of people by machines. The pioneers of automation stoked these fears, by justifying investments in terms of labor reductions. Automation should be seen as sign of progress. It increases effectiveness and efficiency and releases human labor for other tasks. It is key to manufacturing improvement.

Many factory owners and industrial engineers have shared the vision of fully automated factories, without humans – so-called "lights-out factories." Those that have tried to realize it usually ended up with nightmares instead. Fully automated factories are sometimes reported, but they are rarely seen, perhaps with the exception of some parts of process industry plants. Knowing when, where, and how to automate is an essential skill for managers in manufacturing.

This chapter is about automation but it is not a primer on process automation technology, as there are other books for that purpose. Instead, it is about when and where to automate and how humans should work with technology in manufacturing. Whenever human work is considered in the automation effort it tends to be much more successful.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Tell why, where, and when automation is essential in manufacturing.
2. Distinguish between types of automation technology.
3. Use stepwise automation.
4. Understand the role of humans in automation.
5. Apply usability engineering to design human-friendly automation solutions.
6. See the "hidden factory" and know how to improve it with automation.

DOI:10.4324/9781351110310-8

6.1. **Background**

Automata, devices that perform sequences of actions by themselves, have been developed for centuries, mostly for *entertainment*, and in multiple countries. Figure 6.1a–c shows three early examples:

- The Museum of Decorative Arts in Strasbourg, France, holds a mechanical rooster that spreads its feathers and crows with the help of a bellows from the 14th century.
- In 18th century Japan, *karakuri* miniature wind-up dolls served tea.
- About the same time, in France, Jacques de Vaucanson built music playing automata and a mechanical duck.

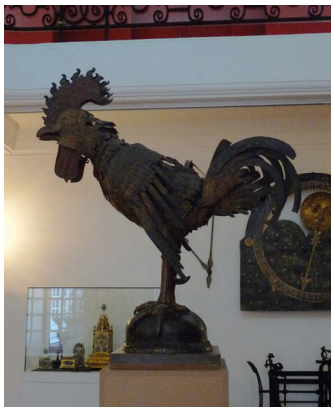
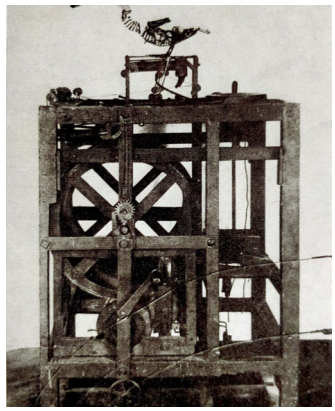
a.  b.  c. 

Figure 6.1 Automatic devices: a) Coq automate (Credit: Ji-Elle), b) Karakuri tea-serving puppet (Credit: PHGCOM), and c) Vaucanson's duck (Credit: unknown/public domain)

To go from toys to manufacturing applications, you needed sequences that resulted in a transformation of materials, the ability to provide outputs at least as consistent as manual work, and planned responses to basic changes in the environment, like running out of materials. The ability to have one device execute multiple sequences of actions was also highly desirable.

These challenges were addressed incrementally over the following 250 years. Today's engineers need to be aware of the older technology because they will encounter it in today's factories. In mature industries, plants have accumulated machines over decades, and multiple vintages coexist.

The Industrial Revolution replaced people, animals, water, and wind as power sources with steam engines. Scottish engineer James Watt developed a centrifugal governor in 1788 that maintained the pressure in a steam engine at a constant level. This invention is still used in household pressure cookers. It was the first device used in manufacturing to keep the operating parameters of a machine within a desired range by *feedback control*, which became central to machine controls after World War II, replacing cranks and levers with servo-motors and relays activated by knobs and buttons. Crank-controlled machines that look like museum pieces, however, are still at work in today's factories.

After his mechanical duck, Vaucanson later turned his attention to manufacturing and is credited with inventing an automatic loom. It was a precursor to the Jacquard loom of 1804, the first machine that could be made to produce different items based on instructions on a punch tape.

In the late 19th century, process industries saw the first automatic plants take in wheat, mill it into flour, pack it into bags, and turn by-products into breakfast cereals. They were dedicated to a process and inflexible.

6.1.1. Early automation in Ford and Toyota

In manufacturing, the integration of sequences of operations within a single machine made steady progress towards taking on more and more previously manual tasks for 150 years without anyone calling it *automation*. The word automation was first used in manufacturing in 1947 by Delmar Harder at Ford. This is what David Hounshell[1] wrote about it:

> *Delmar Harder made in April 1947 a decision to organize an Automation Department within his Manufacturing Division. Harder charged the new unit's engineers and production experts with the development of relatively inexpensive, reliable methods of automating the transfer of materials in process from one operation to another. His instructions were to design and install such automation devices if the cost of each device were less than $3000 (or roughly ten to fifteen percent over the annual wages of an average, fully employed blue collar worker in the automobile industry at the time). Payback for these devices had to be in one year. Thus, for all intents and purposes, to be viable each automation device had to eliminate at least one worker.*

In this context, automation referred to *transfer lines* in engine machining. A transfer line is a manufacturing system which consists of a predetermined sequence of machines connected by an automated materials handling system. Figure 6.2 shows an example of a line with manual materials handling. Transfer lines were designed for working on high volumes of a very small family of parts.

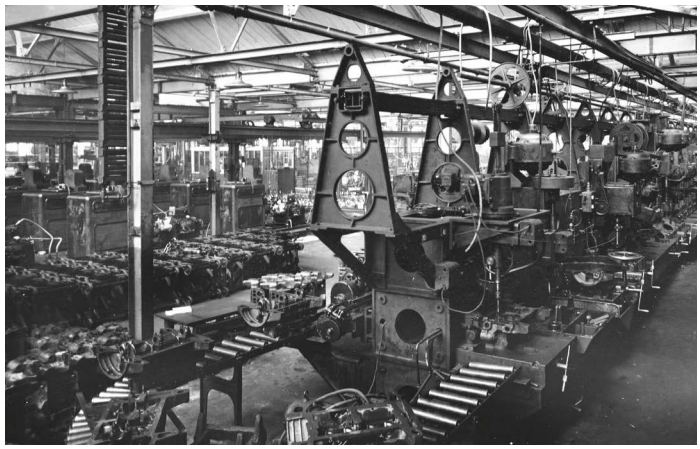


Figure 6.2 Engine production before implementation of an automated transfer line (Credit: unknown/public domain)

John Diebold's 1952 book *Automation*[2] was the first by this title, and Diebold claimed to have coined the word, as a shorter and easier-to-spell alternative to "automatization." He views automation as designating both automatic operations and the art of making things automatic. Diebold saw the field from a

1 Hounshell, D. (1995). Ford automates: Technology and organization in theory and practice. *Business and Economic History, 24*(1), 69.

2 Diebold, J. (1952). *Automation. The advent of the automatic factory*. D. Van Nostrand Company.

broader perspective than Harder. As a term, "Automatization" lives on in psychology, where it refers to the repetition of an action by a human, to the point that he or she can perform it without thinking about it.

At Ford and many other companies, automation, to this day, is *perceived* as focused on eliminating operators. The technical literature on the subject – even in "Industry 4.0" – is nearly mute on the role of people. In German companies, many Industry 4.0 initiatives claim to be centered on people but it is often to promote acceptance with skeptical labor unions. The emphasis on eliminating people not only causes production workers to perceive automation as a threat but it is also impractical because purportedly automatic systems almost always require people to set up, load, launch, and troubleshoot.

Toyota's automation style, known as *jidoka*, is a departure from this approach. The name is an untranslatable Japanese pun. As shown in Figure 6.3, it sounds exactly like the standard Japanese word for automation but is written differently. By adding the radical for "human being" to the character for "move," you get the character for "work."

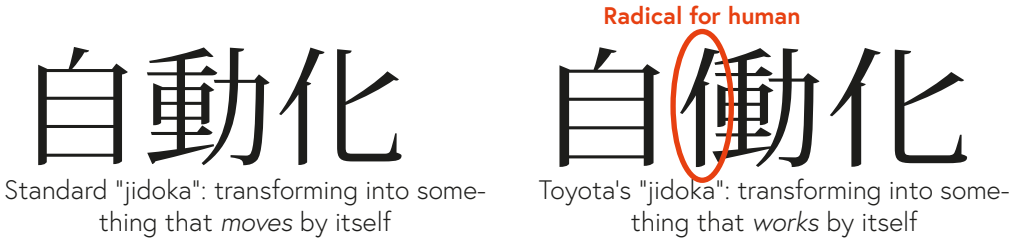


Figure 6.3 Jidoka in everyday Japanese versus Toyota's jidoka

The pronunciation doesn't change but the meaning and the connotation do. Toyota's jidoka *includes the human worker*; the classical jidoka – or "automation" – does not. In its broadest sense, Toyota's jidoka is the engineering of the way people work with machines. In this book, we use the term in this sense (see Box 6.1).

6.1.2. What is automation?

> **Definition**:
> Automation is the elimination of human intervention in a process by means of technology.

While definitions of automation often include mentions of electronic controls, it truly does not matter how it is achieved. Mechanization, for example, designates the replacement of human or animal muscle with machines to perform work. In mechanical processes at least, it is the first step towards automation and was largely achieved in manufacturing by 1950. Beyond mechanization, the focus shifts to telling machines what to do, making sure they do it, making them stop when done, and validating the output, in other words to *control*, which, at one level or another, has been the focus of automation since 1950.

During World War II (1939–1945), the state of the art was turning a gun turret by 63° by rotating a dial instead of using cranks. The challenge was to ensure that the turret would actually rotate by 63°, not 60° and not 70°, and this led to developments in the theory of feedback control that dominated

Box 6.1 **Jidoka in the literature on lean production versus Toyota's jidoka**

In the English-language literature about lean manufacturing, "jidoka" is systematically reduced to just one of its features, the ability of a machine to stop when it malfunctions. If this were all there is to jidoka, it would hardly justify Toyota making it a pillar of its production system on a par with just-in-time.

Jidoka means automation. Regardless of whether it's translated as "automation with a human touch" or "autonomation," it's still a form of automation. Much of the lean literature refers to Sakichi Toyoda's Type G loom stopping when threads broke but that's not all it did (Figure 6.4). It also had *automatic shuttle change*, which solved the problem of what to do when shuttles run out of yarn, which had bedeviled loom engineers for decades.

The goal wasn't switching shuttles but switching bobbins within shuttles, which nobody managed to automate successfully. Sakichi Toyoda was the first to realize that you could prepare shuttles with full bobbins off line, and that switching shuttles was easier than switching bobbins in a flying shuttle.

Sakichi Toyoda's fundamental intent with the Type G loom was allowing one operator to attend to multiple machines. This did require the loom to stop automatically when threads broke but that was only one of many requirements.

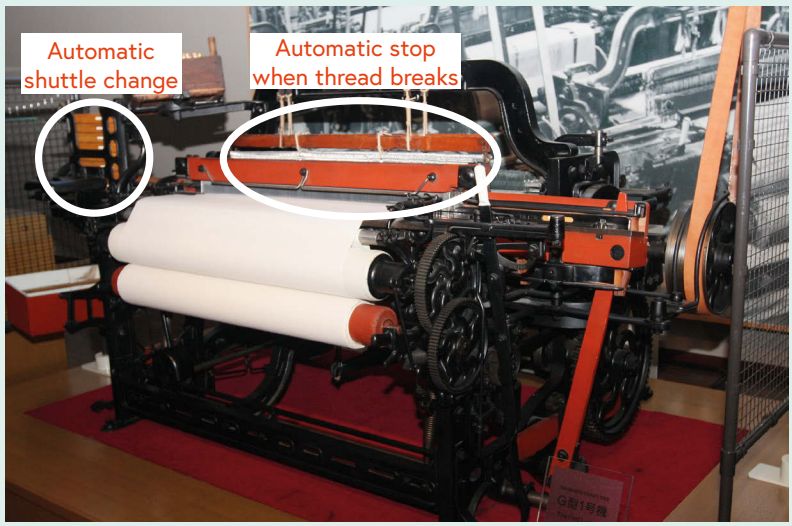


Figure 6.4 Non-stop shuttle change Toyoda automatic loom, Type G, 1924 (Credit: Morio, CC BY-SA)

discussions of automation for decades afterwards. Control theory deals with the control of dynamic systems in engineered processes and machines. Clearly, unless machines follow instructions, nothing else matters.

A textbook on automation from the 1970s would devote most of its space to the math of dynamic systems, proportional–integral–derivative (PID) controllers, stability and compensation, Nyquist

diagrams, and Kalman filters. While these methods are still part of the electrical engineering curriculum, they are not center stage in manufacturing automation because – in mature processes at least – this level of control can be taken for granted. For example:

- Using a stepper motor to turn a turret only tells the motor how many steps to take. To the user of the stepper motor, the control of the angle value of a step is "under the hood" and there is no need to worry about a feedback loop.
- In machine tools, cutters wear out with each workpiece, and the control system adjusts to produce consistent results.
- In sputtering, the metal targets are progressively depleted and the control systems adjusts the process time accordingly to produce a consistent layer.
- In food processing, control systems adjust processes to make consistent products from crops that vary.

Where control theory is still central is in the development of new processes. Otherwise, the main concern in the 2020s is *programming* machines.

Degree of automation

Automation in manufacturing ranges from incremental replacement of one manual operation to radical transformation to a fully automated "lights-out factory":

- To make *incremental* changes, you need to know the details of how work is currently done on the shop floor. As a consequence, it must involve the people who know these details; the production operators. This constrains the company's human resources policy. The hiring of production operators must be with the intent of retaining them for their whole career, and operators who are pulled out of a production team as a result of productivity improvements are not laid off but instead moved to better jobs. All of this requires planning and a rigorous execution discipline. Without it, however, incremental changes do not succeed.
- Radical automation, on the other hand, replaces the process altogether. It should come *on top of a* policy of continuously making incremental changes. Setting up a new line or a new plant is an opportunity to leverage the knowledge accumulated through the incremental changes. If radical automation is pursued on its own, without a simultaneous focus on incremental changes, it is often done with little or no input from the production operators, who fear for their jobs. It rarely succeeds.

Incremental and radical automation should be pursued together. It is not unusual to see two "factories" under the same ceiling at good manufacturers. In a Japanese auto parts plant the authors have seen a fully automated cell with old machines where you could tell that various retrofits, including robots, had gradually taken over the routine work done by operators. Next to this cell was a brand new automatic line that looked different and used no robots, but that incorporated the lessons learned in the old one.

Four degrees of automation can be separated:

1. **Manual**: No automation of the human work task.
2. **Semi-automated**: Automation supporting a human doing the work.
3. **Automated**: Automation replacing a human operator completely for the work, but not for control.
4. **Autonomous**: Automation that fully replaces the human.

Box 6.2 **What machines qualify as automation?**

We often think of any kind of machine as automation. For example, washing machines automate the tedious process of washing clothes and textiles by hand, battery drills automate screwing, and whistle kettles automate boiling water. But if we strictly follow the definition of automation as "the elimination of human intervention in a process by the means of technology" not all machines qualify as automation.

Many processes were first enabled by technological development. In everyday life, airplanes and solar panels, for example, enable processes that humans would not engage in if the technologies did not exist (flying and capturing energy from solar radiation). Replacing a human pilot with an auto-pilot, on the other hand, is automation.

In manufacturing, there are many different machines used for different processes. Other textbooks on machining and manufacturing processes go into detail about this topic. Some examples of special purpose machines include:

- *Shaping:* Injection molding machines, 3D printers, casting machines, etc.
- *Forming:* Hammers, rolling mills, bending machines, brake presses, extruders, etc.
- *Separating:* Cutting presses, laser cutters, eroding machines, etc.
- *Joining:* Machines for welding, soldering, bonding, screwing, gluing, etc.
- *Covering:* Machines for painting, galvanizing, plating, etc.
- *Changing properties:* Ovens, freezers, ionization machines, etc.

Regardless of what a machine does, it is not considered automatic if an operator individually triggers and validates every step it takes. Automating it involves integrating more and more work into each "step," until the complete operation from loading to unloading the work is a single step.

6.1.3. Rationale for automation

The motivation for automation is not the same in every industry. In machining, the main driver is productivity; in nuclear power, the need to protect people from radioactive materials; in semiconductors, there is also the need to protect brittle silicon wafers from breakage and contamination by people. The common thread of the actions that are taken, however, is that they all reduce the amount of human intervention in production.

Hence, any implementation of technology that results in requiring fewer or shorter operator interventions to attain the same results is automation. Conversely, any change that does *not* reduce operator involvement is not automation, even if it includes robotics or computers.

This view of automation is consistent with everyday usage and general English dictionaries. It views automation as a change process rather than a result, and is centered on the objective of this change rather than the means by which it is achieved. An automatic recoil that pulls a hand-tool back into position after use can be part of automation as well as a computer.

Automation and labor costs

Countries with high operator wages have special incentives to reduce operator involvement in manufacturing processes. In 2016, an average Chinese factory worker was paid US$3.60/hour, but the average factory worker in Indonesia or Sri Lanka got only US$ 0.50/hour. The Indonesian worker was around 75 times cheaper than a factory worker in the US, who on average made US$37/hour the same year. Even if Chinese and Indonesian wages are growing at a much faster rate than in the US, it will take many years before they catch up.

The more manual a manufacturing process is, the more severe the competitive handicap of high wages is for that process. Automating the process is a useful response. The most successful manufacturing processes in advanced economies are neither fully manual, nor fully automatic. Instead, they involve interactions between small numbers of highly skilled people and modern machines that account for most of the manufacturing costs and thereby equalize the playing field. The performance differences that can be observed between highly automated factories are usually due to the skills of the people using the machines.

Today, almost every manufacturer in the world has competitors with lower and higher labor costs. The US is a low-wage competitor to Norway but a high-wage competitor to Mexico, which itself is a high-wage competitor to China. In South East Asia, Indonesia, Malaysia, Myanmar, and Vietnam have emerged as low-wage competitors to China. Within China, the Sichuan province is a low-wage competitor to the Pearl River Delta.

Humans and machines

In an internal document, Eiji Toyoda once wrote:[3]

> *"Society has reached the point where one can push a button and be immediately deluged with technical and managerial information. This is all very convenient, of course, but if one is not careful there is a danger of losing the ability to think. We must remember that in the end it is the individual human being who must solve the problems."*

The dark side of the automation vision is a potential massive unemployment of manufacturing workers. But this has not materialized any more than full automation itself. The evolutionary rather than

3 Toyoda, E. (1983). *Creativity, challenge, and courage*. Toyota Motor Corporation, Internal document.

revolutionary nature of this change is only one consideration. Another one is that the jobs eliminated by automation are not the ones that make effective use of human potential.

The only reason manufacturers use people to repeat the same 20 seconds of motion 1,000 times a day is that they do not know any other way. As long as it is done incrementally, the elimination of manual operator jobs through automation is an opportunity for society to make better use of its human potential. Figure 6.5 shows how BMW has automated one of the assembly tasks on a door line.

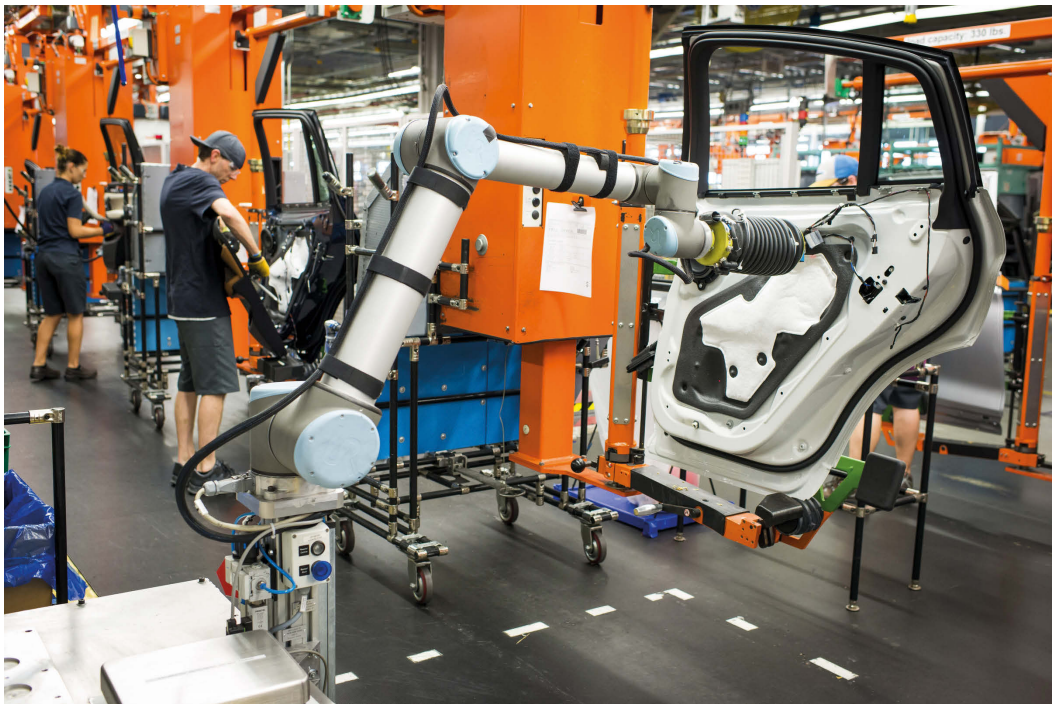


Figure 6.5 Automation and human workers in a BMW factory in Spartanburg, SC, USA (Credit: BMW)

From GM's $40B investment in automation in the 1980s to Tesla's automation woes when ramping up production in 2018 (see Box 6.3), there are many stories of automation failure. Other companies are successful in their automation efforts but it's not in the news, perhaps because they are not pursuing automation per se but automating their processes as part of broader improvement efforts.

At the root of the problems is the confusion between possession of a tool and mastery of it. To make music, an orchestra needs good instruments but it's not enough. The members also need to know how to play. Likewise, a manufacturing plant needs good equipment, but the plant's performance problems are not solved by buying equipment. There is nothing intrinsically wrong with an industrial robot arm, a Flexible Manufacturing System (FMS), or an Automated Storage and Retrieval System (AS/RS) if you know why you are buying them and what it takes to make them work for you.

In almost all countries, there is a shortage of skilled workforce who can operate advanced technology. Such engineers and operators are in high demand, and therefore rare to find and expensive to keep.

Box 6.3 **Automation at Tesla Inc.'s Fremont factory**

In its flagship assembly factory in Fremont, California, Tesla Inc. has been aiming for automation since it purchased the factory in 2010. Today, about 10,000 people work in the 5.3 million square feet factory where Tesla produced 101.312 vehicles in 2017. Twenty years earlier, the same factory was owned by NUMMI, a GM–Toyota joint venture. In 1997, NUMMI produced about 350,000 vehicles with less than 5,000 workers. Although a direct comparison is unfair since Tesla produces more of its parts internally, it suggests inefficient production at Tesla. How come?

The Tesla assembly factory is one of the most automated auto assembly lines in the world. In an interview on *CBS This Morning,* April 13 2018, CEO Elon Musk admitted that technology had come in the way of effective production. "We got complacent about some of the things we thought were our core technology. We put too much technology into the Model 3 all at once; this should have been staged." But the problem at the Fremont factory was not just cars being too technological complex to manufacture, it was also excessive automation of the manufacturing processes. At CBS, Musk gave an example: "We had this crazy, complex network of conveyor belts, and it was not working, so we got rid of that whole thing."



Figure 6.6 Tesla's assembly factory in Fremont, California (Credit: Tesla Inc.)

Automation helps improve productivity, but excessive automation is a recipe for problems. By October 2017, Elon Musk tweeted "We are deep in production Hell." In April 2018, he responded to a question regarding automation in the Tesla factory as follows: "Yes, excessive automation at Tesla was a mistake. To be precise, my mistake. Humans are underrated."

Clearly, Tesla learned this the hard way. By 2019, Tesla has used people to overcome some of the worst problems in the factory and stabilized production at an acceptable production rate closing the gap between production target and actual outcome. The total output in 2018 was 350,000 cars. In 2021, it was 450,000.

In *The Robot Caravan*,[4] Jill Lepore makes fun of the countless authors who, for the past 100 years, have been predicting a future in which robots have eliminated all jobs, manufacturing or not. While Lepore does not go back that far, "robot" is a word from science fiction, specifically Karel Čapek's 1920 play *Rossum's Universal Robots*, in which robots actually kill people.

If robots had a massive impact on jobs, it would show in the employment statistics of the largest economies. As seen in Figure 6.7, it doesn't.

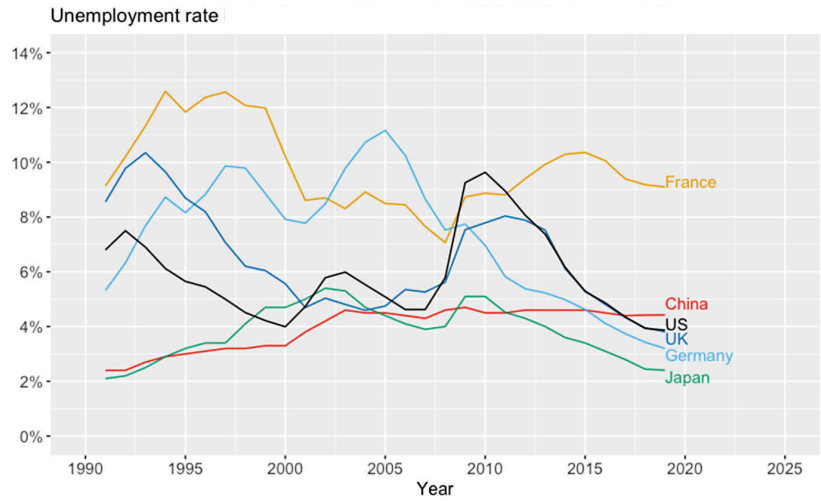


Figure 6.7 Unemployment rate in the six largest economies 1991–2019 (Data source: ILO)

While the curves of different countries are at different levels, they don't show an increase over the past 30 years. Comparisons between the levels of unemployment *among countries* are not as meaningful as they appear because the national statistics bureaus don't define "unemployed" the same way.

The maximum number of hours worked per month to be counted as unemployed are not the same, women's participation in the workforce varies, and the bureaus have varying levels of political independence. On the other hand, we can assume that, in any given country, they calculate them the same way every year, and we are better off comparing their values over time.

During these decades, manufacturers closed factories and laid off workers, disrupting the lives of individuals and communities. And it has rarely been because of robots. Society *as a whole* created more jobs than it destroyed. It remains of course to be questioned if the quality of the jobs – and their wages – were also sustained or improved. What we have been living through is an example of Schumpeter's creative destruction,[5] not a science fiction dystopia.

Stop and Think 6.1
What tasks are automated in your home? What remaining tasks would you like to automate? Why are these tasks not already automated?

4 Lepore, J. (2019, March 4). The robot caravan. *New Yorker*.
URL: https://www.newyorker.com/magazine/2019/03/04/are-robots-competing-for-your-job

5 Schumpeter, J. A. (1942). *Capitalism, socialism, and democracy*, Harper & Brothers.

6.2. Common automation methods

In the next sections, we discuss automation methods that are commonly used in machining, and sometimes in assembly:

1. Transfer lines
2. Rotary indexing machines
3. Flexible Manufacturing Systems (FMS)
4. Chaku-chaku lines
5. Industrial robots
6. Collaborative robots (cobots)

Logistics operations are often automated through as *conveyors*, *automated storage solutions*, or *automatic and autonomous vehicles,* but these technologies are covered in later chapters.

6.2.1. Transfer lines

Transfer lines, as shown in Figure 6.8, have been used for decades in high-volume dedicated production, particularly in the automobile industry, including factories that practice lean manufacturing. Except for being laid out in a straight line for the convenience of automatic materials handling, it behaves like an automated cell; it moves the parts in a fixed sequence through all the machines. The machines are accessible from the back for maintenance, tool replacement, chip disposal, or any other kind of servicing required. In recent transfer lines the straight line layout provides some flexibility, in that you can add, remove, or replace machines but they run in a fixed sequence.

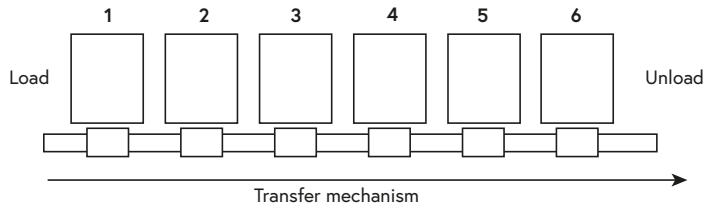


Figure 6.8 Transfer line

6.2.2. Rotary indexing machines

As a concept, a rotary indexing machine (see Figure 6.9) is more similar to a multi-spindle machine than an integration of multiple machines. Functionally, it is like a transfer line, machining parts through a fixed sequence of operations, but with contiguous load and unload stations.

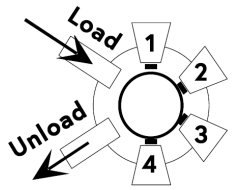


Figure 6.9 Rotary indexing machine

Rotary indexing machines, however, have several drawbacks:

- They are not expandable.
- Access to the work and to the transfer mechanism is difficult for setup and maintenance.

6.2.3. Flexible Manufacturing Systems

Flexible Manufacturing Systems (FMS) are used in machining parts for airplanes, semiconductor production equipment, and other low-volume, high-mix, and tight-tolerance applications. An FMS is comprised of parallel machines with a shuttle and central control system, as shown in Figure 6.10. Note that the FMS is usually fenced in for safety. It is "flexible" because it provides random access to all machines, as opposed to sequential access in transfer lines.

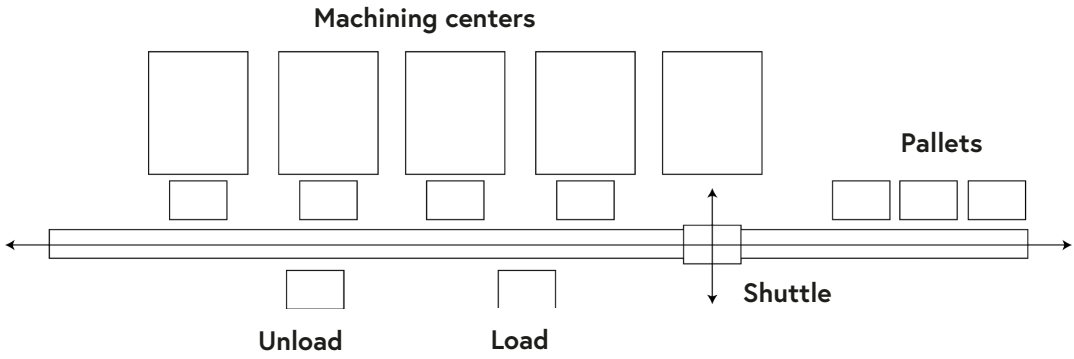


Figure 6.10 Example of an FMS with shuttle

Any part can visit any machine in any sequence, as routed by the control system. This type of system is designed to make hundreds of different items in small volumes, as is required, for example, in aerospace. Even though it is advertised as the best way to satisfy this type of demand, it is by no means a foregone conclusion that it is for a given plant. Simpler machines, arranged in a group of cells, may be able to do better.

A more elaborate example of an FMS is shown in Figure 6.11. The FMS links five machining centers by a shuttle that routes tombstone pallets through one or more of the centers from a raw slab of metal to a machined part.

The loading station is like an assembly station. The fixtures are presented to the operator in vertical slots, in a sequence that matches that of the slabs, and oriented so that the operator can pull them from the slot and slide them on the tombstone without turning them. They are light enough to be hand-carried through this short distance, because most of the required strength and rigidity is provided by the tombstone. The raw slabs arrive sequenced and stored vertically in pushcarts. If they are light enough, the operators can pick them up by hand and mount them onto the fixtures. Otherwise, the preparation of the slab consists of drilling and tapping a tie-down hole at the top, which is then used to attach it to a jib crane at the loading station.

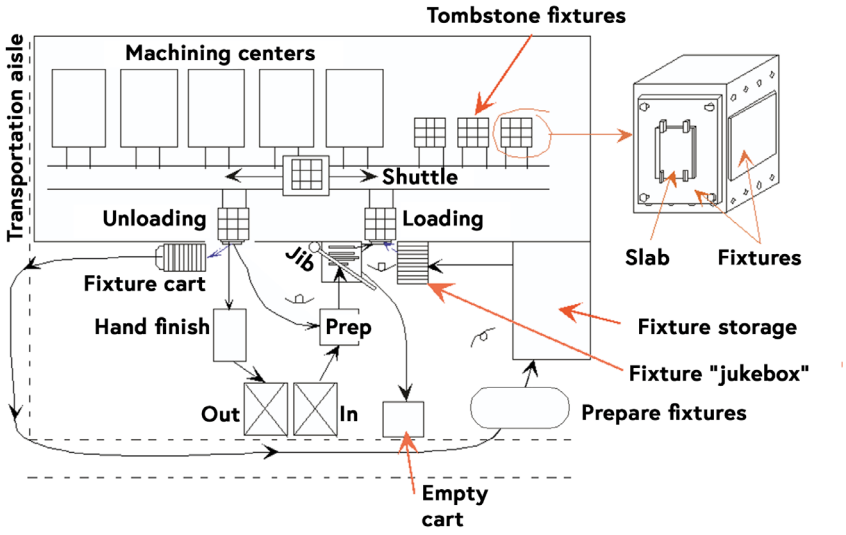


Figure 6.11 Detailed FMS example with fixture flows and tombstone

The machining centers all have the same tooling package, and the central controller routes parts automatically through the machining centers. At the unloading station, the parts are all light enough to be hand-carried. Those that require machining on the opposite side return to the loading station, while the others go to a hand-finishing area, where they may be knocked out of what remains of the slab on the periphery, straightened, and deburred. The fixture flow is designed to avoid interference with the part flow. In this case, the amount of WIP in the automatic system is controlled by the number of tombstones, not the number of fixtures. The main issue with fixtures is to have enough to accommodate the variety of items. There are many more items than tombstone faces, and fixtures are dismounted after each use to make room for different ones.

6.2.4. **Chaku-chaku lines**

Chaku-chaku lines are an advanced form of cells. To appreciate the value of the concept and to be able to implement it, you need the experience of first implementing cells. Chaku-chaku lines are not fully automatic. They are semi-automatic operator-machine systems embodying a philosophy of what is best done by people and by machines. The term "chaku-chaku" literally means "load-load." As a mode of operation, it is opposed to "datsu/chaku-datsu/chaku" which means "unload/load-unload/load".

Figure 6.12 shows the transition from a 1st-generation *machining cell* to a chaku-chaku line. It is clear from the figure that the chaku-chaku line is still a cell, but with the following differences:

- A higher density
- Two buffers removed through the conversion of batch machines to one-piece flow
- Automatic unloading retrofitted to the machines, as-you-go switches that allow the operator to start a machine while walking over to the next one
- Chutes to transfer parts between machines
- Other improvements such as one-touch setups between product

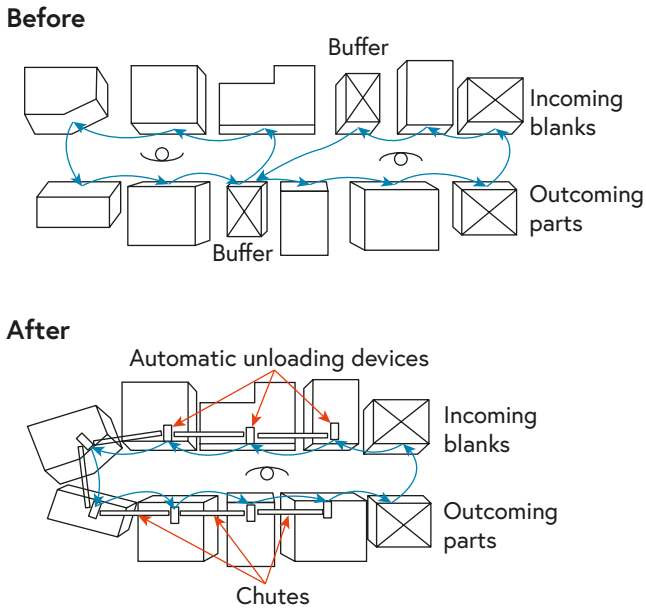


Figure 6.12 Before and after implementing chaku-chaku in a machining cell

While originally a purely machining term, it is now also used to designate assembly cells at a matching level. Figure 6.13 shows an example of a chaku-chaku *assembly line*.

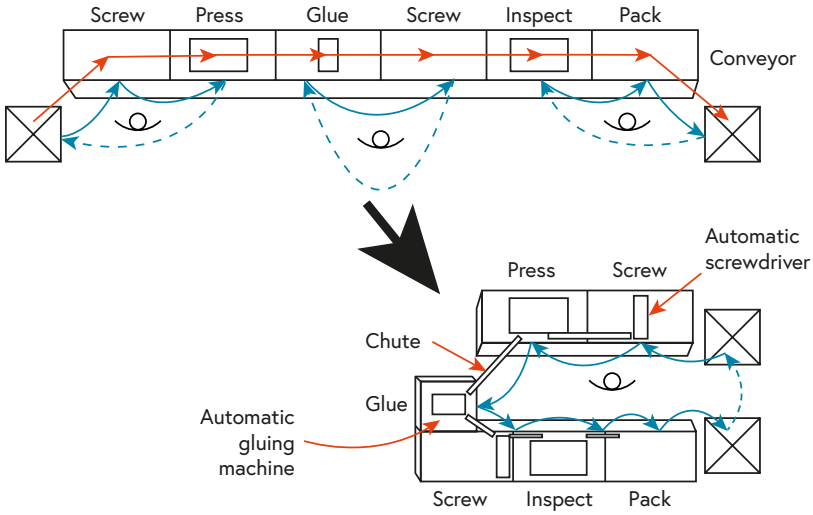


Figure 6.13 A chaku-chaku line for assembly

The transition in the assembly example involves the following:

- Conversion from a straight line to a U-shape
- The removal of a conveyor
- Bringing the assembly stations closer
- Simple automation steps in driving screws and applying glue
- Automatic unloading at each station
- Chutes between stations
- As-you-go switches

- One-piece flow
- One-touch setups

An example of *automatic unloading* is shown in Figure 6.14. It is an example of a simple automatic ejection device retrofitted to a press. Note that the ejection system is powered by the press itself.

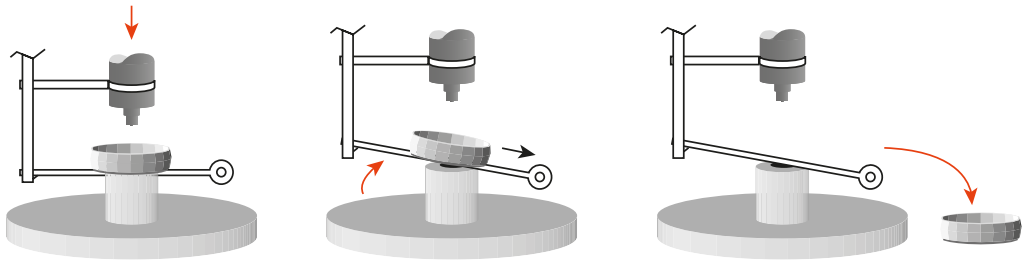


Figure 6.14 Automatic ejection device on press

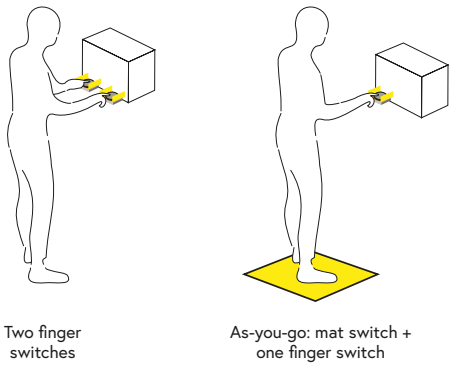


Figure 6.15 As-you-go / nagara switches

As-you-go switches are another characteristic of chaku-chaku lines. As shown in Figure 6.15, the idea of as-you-go switches, also known as "nagara switches," is to allow the operator to start a machine while walking over to the next station, thus combining walk time with machine start time. The machine start operation must of course be as simple as pressing a button, there must be no possible confusion as to the location of the switch, and it must be safe.

The name chaku-chaku line highlights the operator's role in loading parts. As discussed above, this task is, if not automated, at least simplified by work on the fixtures and on loading methods, and by as-you-go switches. We must not, however, forget that the operator has a key role to play in quality assurance, by applying go/no-go gauges to every part after automatic unloading. In fact, in a chaku-chaku line, the operator's work is focused on the product rather than the equipment. Since chaku-chaku lines commonly have a single operator looking after 10 or even 15 machines, many operators get to check the product in every stage of transformation.

Chaku-chaku lines fit within an incremental automation strategy that pursues automatic unloading separately and ahead of automatic loading because it is easier and cheaper. The next level beyond individual chaku-chaku cells is the integration of multiple cells in one big room, with material connections through chutes and flow racks, and operator jobs that involve tasks in more than one cell.

6.2.5. Industrial robots

In popular imagination, automation is often equated with the use of *robots* but they are, in reality, a small part of it. Since Karel Čapek's play, humanoid machines called robots have been in conflict and alliance with humans in numerous novels and movies, including Isaac Asimov's *I, Robot*, Kurt Vonnegut's *Player Piano*, Philip K. Dick's and Ridley Scott's *Blade Runner*, not to mention *Alien*, *Terminator*, or *Star Wars*.

In the 1960s, migration into real life started with kitchen immersion blenders that were marketed as "robots" in spite of being devoid of intelligence. This usage promptly disappeared. Devices like automatic vacuum cleaners don't look like science fiction robots either but they have the intelligence to move around obstacles and cover an entire surface. The public's fascination with fictional robots has encouraged automation providers to call their products "robots" and the press to equate automation and robotics. There are many real products called robot today that are not humanoid and sometimes don't move, blurring the picture.

"Industrial robot" came to designate programmable machines with an arm that can move along two or more axes, which is a far cry from the humanoids of science fiction. The first product called "industrial robot" was introduced by Unimation in 1956 and all it did was move parts between machines. By the 1980s, car bodies were welded and painted by robots. Today, all car manufacturers use industrial robots for body-in-white and paint jobs (Figure 6.16). Millions of industrial robots are at work today.

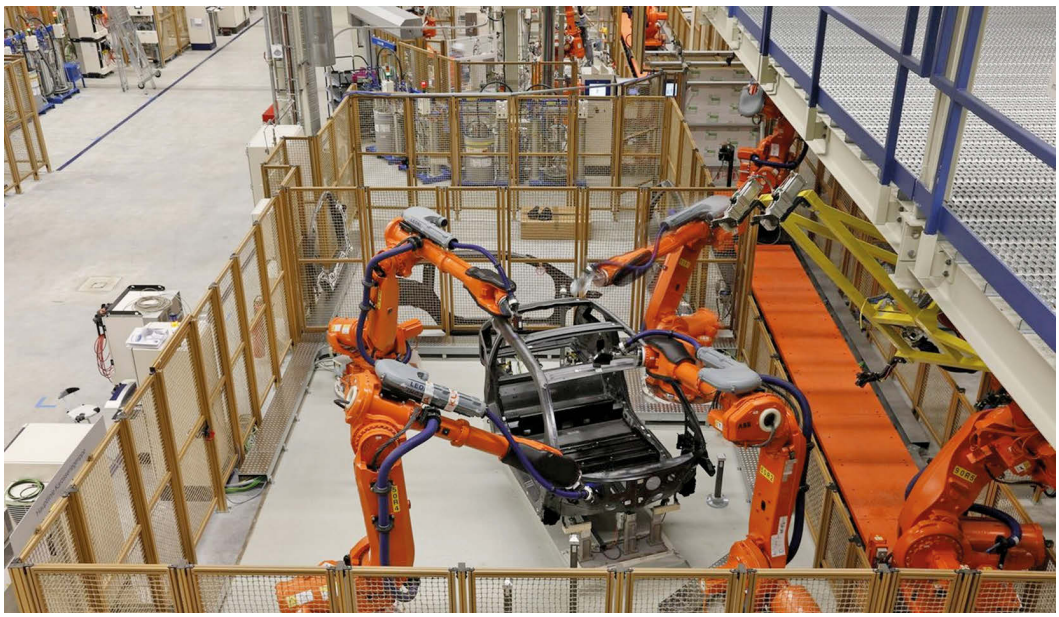


Figure 6.16 Industrial robots at work in a body-in-white car assembly shop

Figure 6.17 shows an example of a fenced-in industrial robot doing final assembly in a washing machine factory. For safety reasons, most industrial robots require physical fencing. Of note, Tesla has less fences in their body-in-white shop in Fremont, CA, than most car manufacturers, which it can because it uses spark-free friction stir spot welding.



Figure 6.17 An industrial robot for final assembly in a washing machine factory

The distinction between a robot and a programmable machine is fuzzy. A CNC machining center or a self-driving truck in a mine are programmable machines that no one calls "robots." On the other hand, the term is routinely applied to the programmable mechanical arms that weld car bodies together and to the autonomous vehicles that move racks of parts in Amazon and Alibaba warehouses.

Furthermore, the "robot" label has lately also been applied to purely software objects that emulate some forms of human behaviors. A useful example is Robotic Process Automation (RPA), which mimics a human's commands in IT systems. Other examples are "robocalls" – phone calls made by machines and addressing recipients in synthetic voices – and systems that place fake updates on social media, so-called "bots."

6.2.6. **Cobots**



Figure 6.18 Cobot working with a human in Skoda (Credit: KUKA AG)

While mechanical arms that go through a fixed sequence of motions to pick and place a part, weld, or paint have been in use at least since 1980, they were, until recently, unable to react to changes in their environment, like the presence of people. They were therefore dangerous to be around. The classical picture of robots in manufacturing is enclosed in cages (see Figure 6.17 and Figure 6.18).

It is now changing with the advent of "cobots" – that is, robots that can share their environment with people. Cobots is short for "collaborative robots." Cobots can work side by side with people without involving any safety or hazard risk to them.

6.3. Principles of automation

In an automation project, following some principles, guidelines, and rules-of-thumb, can avoid common pitfalls. A very basic principle formulated by K. M. Kapp[6] in 1997, is "Understand, Simplify, and Automate," summarized in the quaint but odd acronym "USA." The USA principle is a common-sense approach to automation.

6.3.1. Toyota's 7 steps of automation for machining

Early on, Toyota engineers came up with a list of 7 steps to follow in sequence to automate a machining process successfully. Today, some of Toyota's own processes are so automated already that this list is no longer applicable. Engineers and managers in the many machine shops that are still far from Toyota's level, however, will still find it a valuable guideline. The list is an empirical discovery by Toyota engineers as a result of carrying out many projects. The progression has to do with the tasks carried out by operators, and the list is as follows:

1. Conversion from manual operation to power drives.
2. Conversion from manual feed to power feed.
3. Automatic stop and return to start position at the end of a cut.
4. Automatic unloading of the workpiece, as in chaku-chaku lines.
5. Automatic loading of the workpiece.
6. Automatic error diagnosis.
7. Automatic transportation between operations.

While "automation" commonly refers primarily to automatic transportation between machines, it is often implemented in a context where the machines themselves require human intervention. The 7-steps-of-automation will not let you, for example, use an automatic transportation system to bring a metal part to a workstation where an operator will use a hand tool to tap a hole.

The third step, automatic stop and return to start position, supports semi-automatic operation, where the operator presses "Start" and *walks away*. This step is essential to break the *one-operator-one-machine* model, which is still common in many factories. It can be done on conventional equipment with limit switches and other sensors, with CNC on machine tools, and Programmable Logic Controllers (PLCs) on other types of machines. When first initiated by Taiichi Ohno in Toyota's machine shop in 1950, it allowed hands-free operation by replacing handles with pneumatic/hydraulic chucks, limit switches for end-point detection, and automatic lubrication.

The fourth step, automatic unloading, is the focus of chaku-chaku lines. The key point here is treating unloading and loading as separate automation steps, because unloading is easier and cheaper to automate than loading. Since it does not require precise positioning, it can be done with simple levers and pneumatic cylinders.

6 Kapp, K. M. (1997, June). The USA Principle. *APICS – The Performance Advantage*, 62–66.

Stop and Think 6.2
For machining operations, Toyota automates the transportation between processes only after the automation of the processes. Why?

6.3.2. The 7 steps of automation applied to assembly operations

Do Toyota's 7 step of automation in machining apply to assembly operations? In assembly operations – in contrast to machining – the work itself is not the first target of automation. Indeed, the assembly line may automate transportation between stations at which all the assembly work is done by hand. As discussed in the previous chapter on assembly, this is driven by logistics.

The assembly of mechanical or electronic devices involves the precise positioning of many parts, and automatic devices to do it are particularly difficult to come by in a low-volume-high-mix environment. Automatic assembly works well, for example, for frozen lasagna: robots with suction cups pick and place sheets of pasta, and guns in fixed locations squirt metered amounts of sauce between layers. This is high-volume-low-mix production with loose positioning tolerances.

The sequence of Steps 1 to 6 for machining automation can nonetheless be adapted to assembly. Steps 1 and 2 combine into one: Replacing manual insertion or fastening with hand-held pneumatic or electric power tools. Steps 3 to 6 remain applicable verbatim, and there are chaku-chaku assembly cells. Automatic loading (Step 5) for an assembly operation is an order of magnitude more complex than for machining, because it involves many parts. Automatic *picking*, on the other hand, is much easier, provided each item is supplied in exactly one location and in a controlled orientation. It has the same level of complexity as automatic unloading, because placement on a kit pallet does not require the precision of assembly itself. As a consequence, picking is the focus of automation in flexible assembly lines for, as an example, personal computers.

When we consider manufacturing processes ranging from semiconductors to cosmetics, it is clear that the sequence of steps defined for machining is not universally relevant. At the same time, the idea of having such a sequence as a guideline for managers and engineers is universally appealing.

6.3.3. Scope of in-house development

Equipment modifications as retrofits to conventional machines should be done in-house. This implies that in-house engineers need to be tinkerers, which in turn implies that they must have not only desks in the office area but a lab or engineering "sandbox" where they can try out ideas and show them to operators.

The engineers will then gravitate towards technology they are comfortable with. Mechanical technicians may, for example, distrust electronics and choose to use fluidic switches that use air flow to turn devices on or off. The resulting devices bear this kind of "signature" from the people who developed them, and different technologies may be used in different plants for the same function.

How far should in-house development go? New industries have no equipment suppliers, and therefore offer no alternative to in-house development. An in-house built, special purpose machine can be a source of considerable competitive advantage, as in the example of protective gloves in Figure 6.19.



Figure 6.19 Many special-purpose machines must be built and maintained in-house (Credit: Song Qiuju/123RF)

This is especially true if its inner workings are complicated and can be protected. The dominant semiconductor companies of the 1960s, Fairchild and Texas Instruments, both made their own process equipment, but no semiconductor company would attempt it today, as their homegrown machines would be unable to match the technology offered by the now robust equipment industry.

What needs to be done in-house, however, is the customization of the standard machines bought off the shelf to the specific uses they are put to in the plant, as well as the complete development of equipment for secondary operations that cannot be bought. Possession of equipment does not confer a competitive advantage. You also need to master the art of using it.

6.3.4. Centralized versus decentralized control

The approach described in the "seven steps" is implemented bottom-up. An increment of automation is undertaken, for example, as part of an effort to reduce the number of operator positions in a cell by one. If a cell needs 4 operators, 3 of which are fully utilized and the other 1 is utilized 20% of the time, then we can go down to 3 operators by reducing the total labor time required per part by 6.7%, and an increment of automation on one of the machines in the cell is one of the ways it can be done.

The selection of the target machine is therefore the result of a top-down analysis, but the actual automation is bottom-up, machine by machine. As a result, each machine has its own controller, and, until the level of automatic transfers is reached, the coordination between machines is done by people. When we get to the stage of automatic transfer of parts between two machines, then the controllers of the upstream and downstream machines need to have a "conversation" to establish that the downstream machine is ready before the upstream machine releases a part to the transfer mechanism.

Local handshakes between every pair of machines in the cell are then sufficient to coordinate its operations. In other words, if you automate a cell, machine by machine, and step by step, you obtain an automatic production system that does not require a central controller to coordinate machines and materials handling systems. Central supervisory control may be useful for other purposes, such as equipment monitoring/alarm generation, but not to trigger part movements in normal operations.

This is different in concept, and much easier to build and maintain, than the centralized controllers that coordinate the materials handling system with machines in systems ranging from transfer lines to FMSs sold by the machine-tool industry.

6.3.5. Automation levels and skills management

When a factory is built, it is usually with an equipment set that is consistent not only in terms of process capabilities but also in its level of automation. It rarely stays that way, and a mature plant most often has a mixture of equipment vintages coexisting and sometimes producing the same parts with different processes.

This raises the question of whether the level of automation of all the machines in a cell or a line should be consistent. It is certainly not a requirement. It is not uncommon to see a lathe or a mill operating at level 3 in Toyota's 7 steps, with automatic stop and return to start position, followed by manual deburring or hand finishing.

Inconsistencies in automation levels within the shop floor, however, can cause *organizational* problems, due to the differences in tasks and human skill requirements of the different kinds of machines. Machinists who run conventional equipment may not be computer-literate, while those who run CNC machines may not know how to tell what is happening to the machine from the way it sounds, smells, or vibrates. The skillset of operators involved with the more modern, more automated equipment is not only different, it is also rewarded with higher wages, which may be perceived as a two-tier system.

The step-by-step approach brings along the workforce level by level. This not only avoids having a two-tier system but also makes the operators better able to contribute to the effort.

6.4. Working with machines

When automating processes, new technology changes the work people do but does not eliminate the need for people. How people work with technology determines the performance of manufacturing systems, and has since the Industrial Revolution.

6.4.1. How much technology must operators understand?

A chef can have excellent egg-cooking skills without understanding the chemistry of coagulation. He or she knows how long it takes to boil it hard or soft, but doesn't need to know what happens within the white at a molecular level, because it wouldn't help cook better.

Between a basic principle and a physico-chemical theory, there is a range of levels of understanding of a process that can be formulated. How deep a machine operator needs to go is a function of native ability, level of education, and relevance to job requirements. Although managers and engineers rarely give them credit for it, most shop floor operators have the native ability to understand manufacturing processes at a sufficient level to contribute to improvements. Conveying the information effectively may be a challenge, but it can be met, and operators as the ones in Figure 6.20 are a receptive audience.




Figure 6.20 Some knowledge into the inner working of machines help troubleshoot and run them better

While the same native abilities can be assumed in every factory, the level of education of the workforce varies both with the social environment and the company's human resource policies. The Detroit area workforce of the early 1900s, for example, included many immigrants from the farmland of multiple European countries, with at most a primary school education in a language other than English. Not surprisingly, such an environment made for simple jobs, with expectations limited to the repeated execution of a few moves. Nine decades later, in Georgetown, KY, Toyota was hiring college graduates as production workers, with expectations including not only execution but also some level of troubleshooting and improvement, all of which call for a deeper level of process understanding.

The relevance of topics has to be analyzed case by case. Traditional aerospace machining, for example, includes cutting processes that remove 90% of the metal from a workpiece, and aerospace machinists receive training in tool feeds and speeds. This topic is less important in automotive engine block machining, which instead involves a large number of small cuts on castings that are already near their final shapes. Because of its immediate relationship to their activities, all machinists benefit from knowledge of the alloys the workpieces are made of, the life of cutting tools, and the shape of the chips produced when cutting.

The relevance of knowledge also depends on the focus of improvement. If the process is new technology or mature technology applied to a new product, then the first focus of improvement is often either the capability of the process or its capacity. If the process cannot reliably hold the required tolerances, is too slow for the expected production volume, or consumes too much raw materials, these issues must be addressed by teams that combine knowledge of process technology and problem solving.

In such teams, operators typically play a marginal role – if any. Eliminating leaks in die casting, for example, requires tweaking multiple parameters of the process. To have an idea of what may work, you need to know the physics of die casting; to find the "sweet spot" in the parameter space quickly,

sometimes with statistical Design of Experiments. Cutting the milling time on a wing component by a factor of three may require starting from different forgings, changing cutters, and rewriting NC programs.

This is the work of engineers and technicians, not operators. Once the process is capable and has the needed capacity, its numerical parameters are best handled by the control system, letting the operator deal only with discrete changes, such as startup/shutdown, setup, operation start, unload/load, or malfunctions identified by parameters out of range or defective output. The focus of improvement then shifts to the flow of materials and the work of operators, where most projects do not require process knowledge beyond the times required for manual and machine-driven activities. This includes collocating machines, arranging for one-piece flow, and even most of the setup time reduction and mistake proofing.

6.4.2. **Alienation**

As consumers, we routinely use machines without understanding their inner workings. We use computers, microwave ovens, ATMs, smartphones, digital cameras, TVs, and many other devices and neither know nor care to know how they work. Many machines used in the workplace are the same way. You place a document in the feeder tray of a copier, press "Start" and collect the copies. You chuck a blank in a CNC lathe, press "Start" and collect the turned workpiece. For the sole purpose of getting output from a machine in normal conditions, you don't need to know what happens inside.

From the point of view of human understandability, there are three types of machines:

- **The principle of the process is visible.** A machinist sees the cutting tool attacking the workpiece and chips flying out, or the brake press bending metal. Likewise, extrusion does not appear mysterious to anyone who has squeezed the toothpaste out of a tube. Molding is another example.
- **The change in the state of the workpiece is visible.** For example, workpieces come out of sputtering or anodizing with a different color. The process itself is usually not visible, and, unless specifically taught, the operators will not understand its principle, but they can easily tell whether a part has been through it or not.
- **There is no visible change in the appearance of the workpiece.** Some processes affect the internal structure of materials without leaving any trace on the outside. Annealing, for example, relieves the stresses introduced in a piece of metal by machining or forging, but there is no change on the outside that the naked eye can see. The same is true of ion implantation in semiconductors: a dopant is inserted into the crystal structure of a silicon wafer by bombarding it with a scanning beam of accelerated ions, changing its electrical properties but not its appearance.

In principle, only process engineers need to know the physics, chemistry, biology, and/or data science of what machines do. In reality, however, process understanding makes a difference in the operators' ability to run machines. Only an operator with some knowledge of the processes carried out inside the machine can do the following:

- Participate in setups
- Monitor tooling
- Help diagnose failures
- Contribute to improvements in process capability

Not understanding how household appliances work is a fact of modern-day life that does not particularly concern sociologists or philosophers. Work, however, is different, because operators interact with machines for an entire shift every day for years and decades, and it is their livelihood. The loss of control associated with the transition from handicraft to industry has inspired a philosopher like Karl Marx to describe it as "alienation." In the 1960s, Robert Blauner[7] argued that alienation in industrial production exists in four ways:

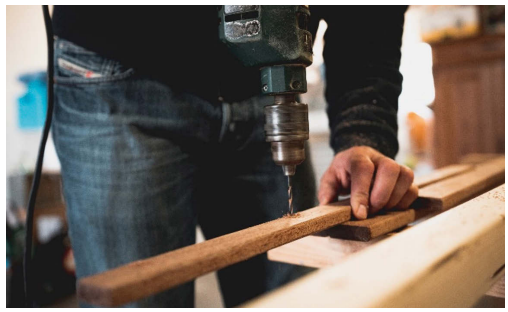
- **Powerlessness.** Workers are not in control of their work process or activity.
- **Meaninglessness.** Workers do not see the function, purpose, or meaning of their activity in the context of the whole organization.
- **Isolation.** Workers do not see themselves as part of a social group or community in the workplace.
- **Self-estrangement.** Workers are not involved in their work as a means for self-expression.

Figure 6.21 shows increasing degrees of alienation:

1. Using hand tools to form a piece of material is a degree of alienation from forming it with bare hands.
2. Using powered tools adds another degree because the human is no longer controlling the force.
3. Computer controls like the CNC machine shown in Figure 6.21 represent a 3rd degree of alienation because they execute and validate an entire sequence of steps without human intervention.
4. Supervisory control is a 4th degree of alienation because the human is not even by the side of the machine.



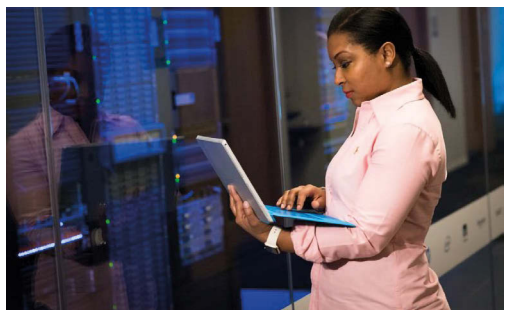
1st degree of alienation



2nd degree of alienation



3rd degree of alienation



4th degree of alienation

Figure 6.21 Increasing degrees of alienation from manufacturing processes

It is also possible to envision a 5th degree where artificial intelligence (AI) completely removes the human from the manufacturing processes. Those who have attempted this to date have found that the available technology was not capable of it.

7 Blauner, R. (1964). *Alienation and freedom: The factory worker and his industry*. University of Chicago Press.

6.4.3. **Unloading and loading workpieces**

Operators' work with machines usually involves a role in unloading and loading workpieces. Because of the great variety of sizes, weights, positioning, and cleanliness requirements, there are only a few general principles that can be stated about these activities. Unloading is easier than loading and is normally automated first. Ejecting a part is easier than placing and seating it properly in a fixture or chuck. When the workpiece moves through a sequence of machines, the requirements on unloading are as follows:

- Preserve the first-in-first-out (FIFO) sequence of parts for traceability of quality problems.
- Keep the workpiece oriented for easy loading at the next machine.

In an example of injection molding of donut-shaped parts, Figure 6.22 shows two different ways to eject parts. The upper process is the common approach of dumping the parts into a bin. There are numerous problems with this approach. It does not orient the part, hinders part counting, collects dust and dirt, and forces the last-in-first-out (LIFO) principle. The lower process shows a better design.

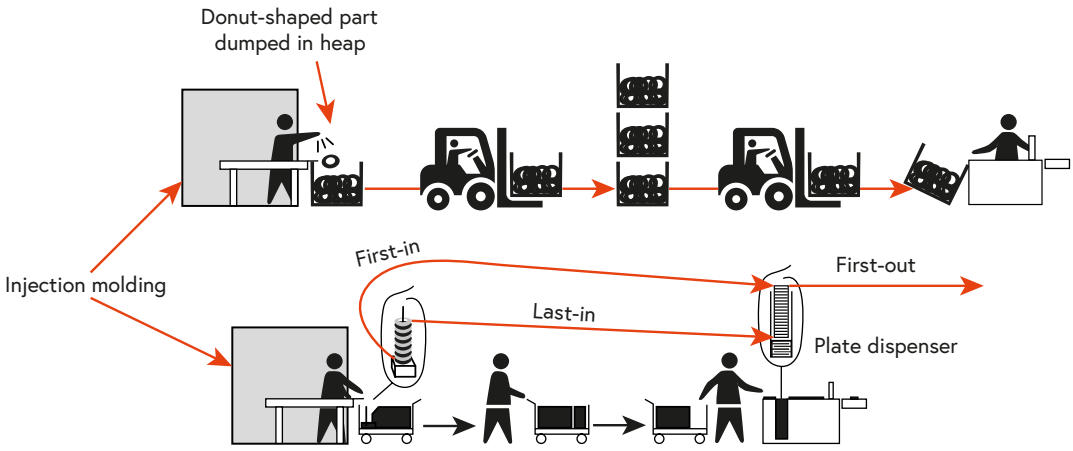


Figure 6.22 Example of unloading from an injection-molding machine

The unloading and loading of parts that are too heavy to be hand carried is done by default with cranes, which is inefficient for the following reasons:

- Workpieces often need to be modified or placed in a harness to be lifted by a crane.
- Positioning a heavy workpiece in a fixture while dangling on a crane is difficult and can be unsafe.
- When bridge cranes are used, they are always too few, and work is delayed waiting for a crane to be available.

A preferred solution, whenever possible, is to use devices that hold the workpieces from below and move them horizontally between machines arranged to have the same work height.

Many machines have pallet or fixture exchangers that allow operators to unload a finished piece and load a new blank while the machine is processing another one, so that the machine can run continuously. Several pallet exchanger concepts are shown in Figure 6.23. The shuttle type requires the operator to work on both sides of the machine, and should only be used for very large workpieces. For smaller workpieces, the rotary type is preferred, because it allows the operator to do the entire unload/load step from the same side.

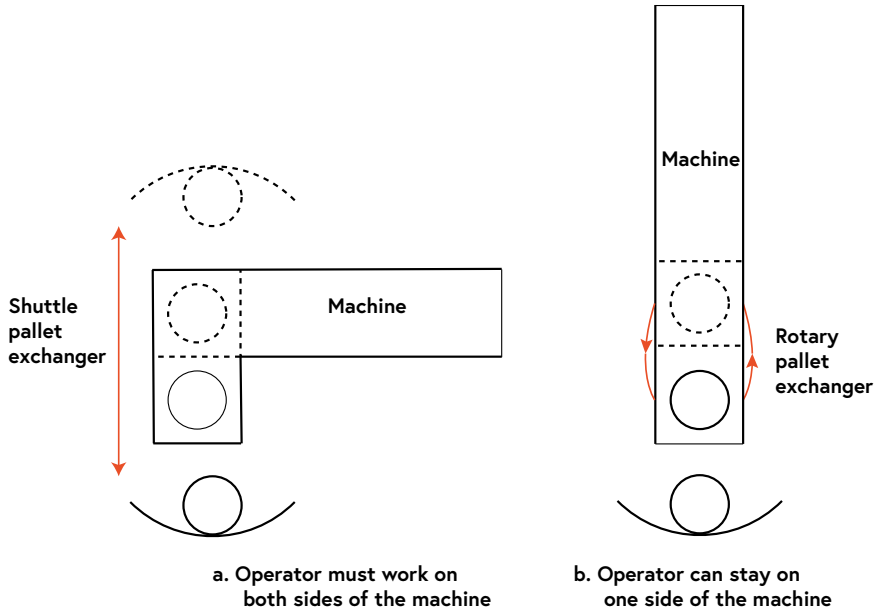


Figure 6.23 Two different pallet exchanger concepts

6.4.4. Human-Machine Interfaces

Production machines used to be controlled through cranks, wheels, valves, and dials that were all specific to each type of machine. Today, they have been replaced with panels of buttons, indicator lights, and screens, and there are many techniques that can be applied to ensure that, if you intend to use one of these controls, your hand does not reach the wrong one. The interaction between a touch panel or physical knobs, buttons, joysticks, and switches and the human operators is known as the Human-Machine Interface (HMI).

A Supervisory Control and Data Acquisition (SCADA) system provides a single, clear human interface to multiple machines. It improves the interfaces of multimillion-dollar automated system at a cost in the thousands of dollars. It allows one operator to monitor multiple machines while doing other work, without any loss in response time.

Machines come from suppliers with their own interfaces:

- A panel of push buttons, knobs, joysticks, and other controls
- Lights, dials, and other discrete indicators
- A screen of status information
- A keyboard for editing, downloading, or referencing programs

These generic interfaces (see Figure 6.24) make it possible to operate the machines, but rarely fully meet the needs of a specific factory. Indicators may be distributed around the machine, with, for example, a thermometer in the front and a pressure gauge in the back, making it difficult for the operator to notice a simultaneous excursion of temperature and pressure. All the parameters should be visible from one single location.



Figure 6.24 The Human-Machine Interface is the link between the machines and operators

On the other hand, some controls should be accessible from anywhere around the machine. If the operator, while working on the machine away from the control station, notices the need for an emergency stop, it should be doable without running back to the big red button on the control panel. This can be done by using remote controls or stop ropes. Routine operations should not require a keyboard. Whatever IDs the controller needs to have from workpieces or tooling can be entered faster and more reliably by scanning barcodes, QR codes, or RFID tags. Keyboards are in fact only required for machines that are programmed directly on their controller, and then not by operators but by programmers. The most common solution is for programmers to develop programs offline and download them to the machine controllers.

Unfortunately HMIs are more often than not confusing. They are also often inconsistent across machines of different types that must be run by the same operators. Machine users, however, do not have to live with flawed interfaces. They can retrofit their own to embedded machine controllers, through such tools as external operator terminals for Programmable Logic Controllers (PLCs), or SCADA software running on an external workstation networked with the embedded controller. It is feasible at costs that most manufacturing organizations can afford, *if they know what they want.*

The panels of flashing lights and buttons of science-fiction spaceships is not the ideal to strive for. The ideal control panel instead has a "Start" button and nothing else. All other controls and indicators are countermeasures to the imperfections of the machine

Stop and Think 6.3
Why are operating panels for machines different from each other?

6.4.5. Monitoring and alarming

Much of the work operators do on automatic systems after loading them is monitoring for anomalies, restarting them after micro-stoppages, and calling for assistance in case of major failures. Many automatic systems have control rooms, cabs, or just workstations with operators assigned to them whose sole job is problem detection. These operators experience the same problems as locomotive drivers and airplane pilots on automatic: staying awake while nothing happens for extended periods of time, and then acting promptly and appropriately when something does.

To maintain pilot alertness, to this day, airlines have not come up with much more than coffee and copilots. On a shop floor, however, it is possible to combine the work of monitoring the equipment with physical activity that keeps the operator awake by requiring movement. As long as problem detection is not automatic and the equipment requires constant monitoring, then an operator who moves parts between machines will be more effective *at monitoring* than one who has nothing else to do.

To assist the monitoring task, manufacturers can use simple techniques that make anomalies more apparent. Three examples are shown in Figure 6.25. The first photo is a simple fan added to a cooling fan; operators will pick up differences in the fan's whistling sound. The second photo shows a white open filter marked with a red-painted mounting ring, where dirt will be easy to spot. The third photo shows the use of red and green visual labeling to more easily see when a parameter is out of spec. A fourth method, which is not shown below, is to replace all machine covers with transparent thermoplastic covers, which enables direct visual control of the machines' inner workings.

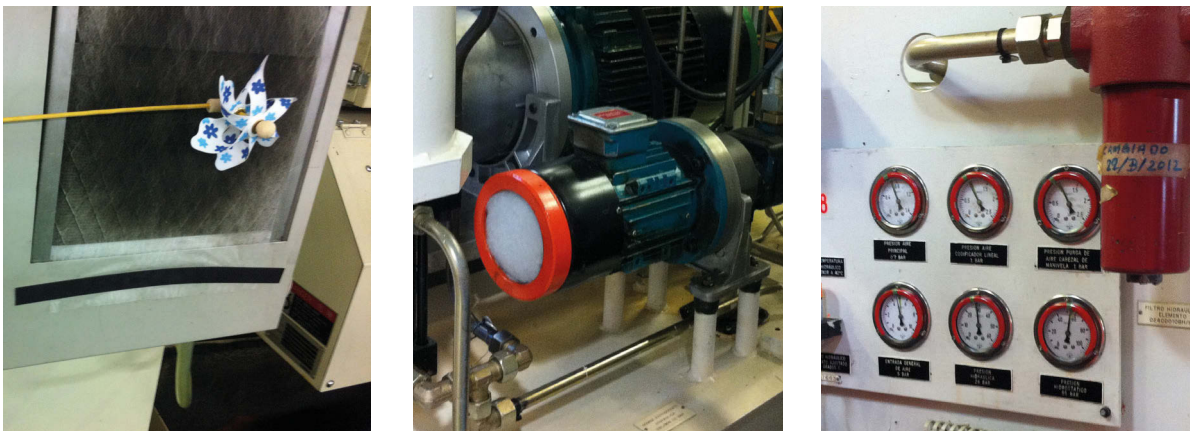


Figure 6.25 Simple retrofits to increase operator alertness

On a machine tool, the problems that require operator intervention include tool breakage, overheating, coolant system failure, chip accumulation, and others that can be categorized and listed. For each of them, the engineers need to design ways for the computer controls to detect them through changes in torque, temperatures, or fluid pressure. There is no doubt that sensors can be both more sensitive and more reliable at this task than people.

In 2022, it is straightforward and cost-efficient to implement anomaly detection sensors and software that improves its performance by learning as it accumulates a history of responses. These can be local solutions or interact with online databases in the cloud. Companies such as Amazon offer out-of-the-box IoT retrofits for this purpose. In this view, photos such as the one to the left in Figure 6.25 could be seen as quaint and obsolete. They are, but such makeshift, cheap solutions pave the way to the development of AI-driven control and maintenance.

When sensors detect a problem, the next challenges for the control system are:

1. To shut down the machine gracefully – that is, without damaging the parts in process and in such a way that the process can be restarted where it left off.
2. To issue alarms that draw the attention of the operator and possibly of the Maintenance department.
3. To communicate the condition of the machine to the operator in the most accurate and easily understandable fashion.

None of the above is easy. In machining, a CNC can remember where it stopped and resume there. If the process involves chemistry, on the other hand, it is much less obvious that it can be stopped and restarted. If the process involves a metal smelter in heavy industries, stopping and cooling it down is not an option if it holds liquid metal. In general, exception handling in the execution of computer programs is still the subject of active research and does not have a generally accepted solution.

6.4.6. Communicating alarms

Alarms can be sound or lights, or both. An alarm sound should be loud enough to get the attention of the people who must respond but no louder. If a building evacuation is needed, the corresponding siren is used. On the other hand, we only need to get an operator who is currently presetting tools to come and look at the controller screen, then a buzzing sound no louder than an alarm clock should be sufficient.

A common device used to communicate equipment status to others is the tower light as shown in Figure 6.26. Even though tower lights are simple devices, it is possible and even common to implement them in a way that makes them close to useless.

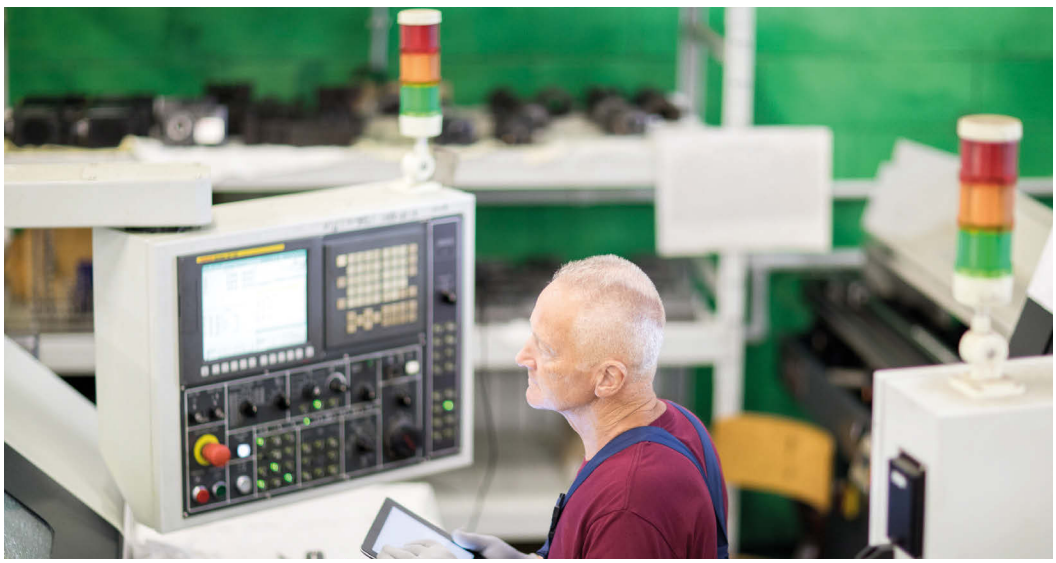


Figure 6.26 Typical machine tower lights

A recent development in manufacturing is to communicate alarms via smartphones or smartwatches. One advantage is that it will reach operators who are physically distant and it can already give an idea of the severity of the alarm. It can be an advantage when operators run many machines in parallel, but it also introduces a new technology with its own challenges.

6.4.7. Usability engineering principles

> **Definition**:
> Usability engineering concerns the design of human–machine interfaces that have high user friendliness.

Machines should be designed so that they are intuitive and easy to use by operators. Usability engineering applied to the HMIs of production machines is key to reducing training costs, enabling operators to become multiskilled, and preventing mistakes with modern, computer-controlled machines.

To apply usability engineering, a manufacturing organization needs control over the machines' operator interfaces, either by requiring suppliers to meet specs, by hiring a system integrator to reengineer the suppliers' default interfaces, or by developing the interfaces in house.

The foundation is the individual controls, and the way control panels ensure that operators press the buttons corresponding to the actions they want to take. The next step is the combination of actions through which the operator causes a machine to process a part or a load of parts through an operation.

This level of control can practically be achieved only with programmable machines. Programmability is the most sweeping change affecting human-machine interactions in the past 70 years. It has changed the jobs of operators, maintenance technicians, and manufacturing engineers, and caused the emergence of control programs as a new class of assets.

The following paragraphs focus on eight principles to use in operator interface design:

1. Affordance
2. Visibility and labeling
3. Feedback
4. Distinctiveness
5. Consistency
6. Arrangement by frequency
7. Cultural constraints
8. Semantic constraints

Affordance

In *Design of Everyday Things*, Don Norman[8] defines the *affordance* of a device as the use suggested by its shape. Through their affordances, hooks are telling us "pull me" or "hang something on me"; plates, "push me"; knobs, "turn me"; scissors, "put your fingers into the finger rings." Accordingly, equipment is easier to learn and less error-prone to operate when the buttons, joysticks, levers, hoops, and other devices in the control panel work as their affordances suggest.

Visibility and labeling

The need for clear labeling of controls is obvious, but there are ways this clarity can be lost. For example, the use of indirect labeling is a common error. Since engineering drawings have each datum identified by a number in an call-out, with descriptive names and tolerances attached to this number in a table in a corner of the drawing, an engineer may fail to see anything wrong with the use of the same approach on a control panel, as shown in Figure 6.27.

8 Norman, D. (2013). *The design of everyday things*. Revised and expanded edition. Basic books.

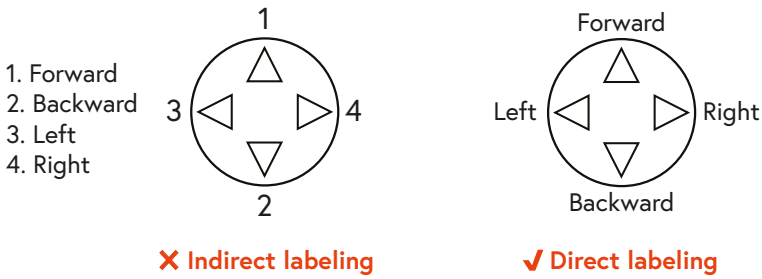


Figure 6.27 Direct versus indirect labeling

Feedback

Through its controller, the machine must let the operator know when an event has occurred. It must acknowledge an operator action on a control and signal changes in the condition of the equipment, and it must do it promptly. Feedback that is less than instantaneous confuses the operator as to whether the action actually took place, and often causes him or her to repeat it. Feedback can be provided through multiple senses:

- The button or key clicks when pressed.
- Lights go on, change colors, or start flashing.
- Messages appear on a screen.
- The machine or the floor starts vibrating.
- The machine emits startup or shutdown noises.
- A buzzer goes off.
- A machine-specific tune is played on the public address system.
- An abnormal noise or smell signals a malfunction.
- The machine perceptibly heats up or starts.

Distinctiveness

Each control must look and feel different from the others. The look provides visibility, but touch can also play a role. Rows of identical buttons are errors waiting to happen. The beer keg handles are attached to two otherwise identical joysticks in the control room of a nuclear power plant and serve to make them not only look different but also feel different to the touch, thus reducing the risk that an operator will use the wrong one even when looking in another direction.

Consistency

Variations in control panel design between machines increase training costs and error rates among operators in charge of multiple machines. Consistency makes it easier for operators to switch between machines.

Arrangement by frequency of use

Some actions are taken by operators with every part or machine load, but operators do not usually start up and shut down the machine with every part. In addition, there are also actions that are only taken by maintenance technicians. These different types of controls should ideally be separated and infrequently used controls protected to avoid accidental activation.

Cultural constraints

Another way to make controls easy to learn and use without errors is to take advantage of conventions that are already known to operators from daily life, such as QWERTY keyboards, traffic Signs. You can also use the colors in traffic lights, for hot and cold water on faucets, or for boy and girl baby clothes.

Semantic constraints

The label attached to a control should not clash with its meaning. No engineer in his right mind would call "open" a button that closes a valve but, to shutdown a Windows computer, the user needs to click "Start," which confuses every new user. The designer's logic is that "Shutdown" is a program, that programs are launched on the "Start" menu, and that it is therefore logical to invoke "Start" in order to shut the machine down. This example shows how violations of semantic constraints occur when the engineer's logic differs from the user's.

> **Stop and Think 6.4**
> Consider the eight usability principles in relation to a machine you use in your daily life – for example, a coffee machine or a washing machine. How could designers have improved the usability?

6.5. The hidden factory

The first thing engineers think about in factory design are the materials that go into the product and the processes that directly transforms input to output. The indirect processes often get much less attention. Sometimes the importance of indirect processes are first realized when the machine does not run or the right tools are missing. Therefore, we call these processes the "hidden factory." It involves tool logistics, waste handling, and recirculating devices.

6.5.1. Routine operations with automated systems

Routine operations on automatic systems still involve operators in mounting and dismounting parts, rotating fixtures, pallets, or molds into and out of machines, activating the right control software, and monitoring tools.

In principle, as soon as we reach the level of automatic loading the operator should not have to touch the machine as it goes through multiple cycles. Machines that are described as having automatic load/unload capabilities, however, frequently have *pallet exchangers* that shuttle pallets in and out, while parts are mounted and dismounted on these pallets manually. This is different from having the machine pick blanks and return finished parts in containers that are ready to move on. The pallet exchangers do not really automate loading and unloading, but decouple these tasks from the operation, so that the machine can work on other parts while operators load and unload offline. It eliminates waste in machine time but does not necessarily save labor.

If the part stays attached to the same fixture on the same pallet through multiple pieces of equipment, as, for example, in a transfer line, then all the intermediate operations between the first and the last stations are truly hands-free. If, on the other hand, a part requires multiple lays on different fixtures, it may go through manual mounting and dismounting several times at the same machine.

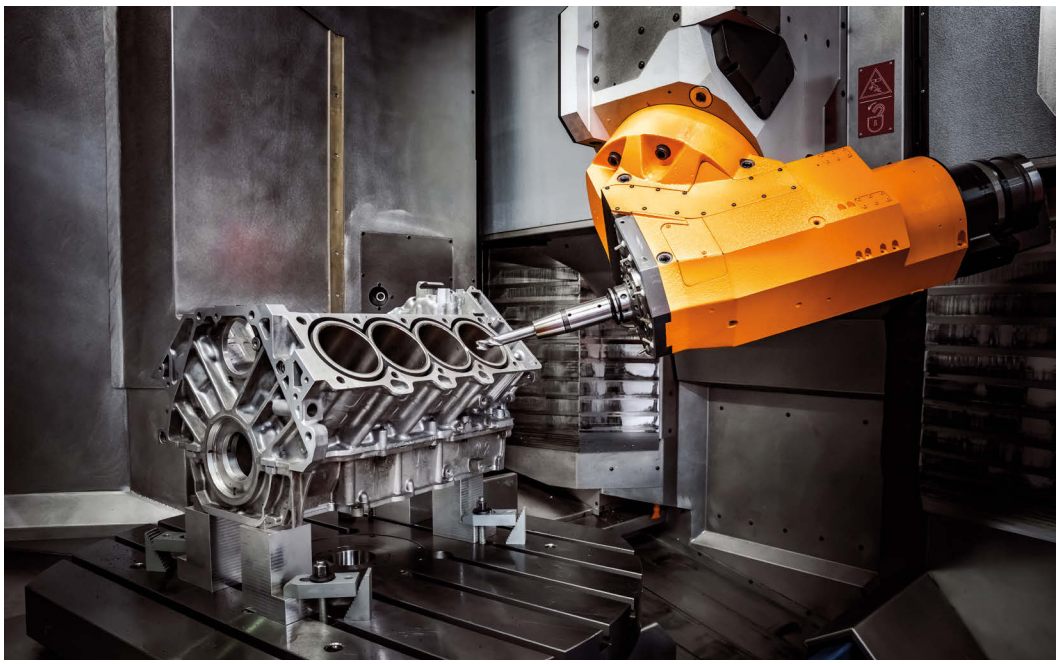


Figure 6.28 The engine block is manually mounted to a fixture that automatically enters the CNC milling machine

6.5.2. **Recirculating devices**

Parts go through the system once or a few times and then leave. Fixtures and pallets, on the other hand, keep recirculating through the system, which has the following consequences:

- The work done on the fixtures, pallets, molds or other recirculating devices must not be allowed to slow down the progress of parts. For example, parts should not wait while operators clean and check these devices. If need be, a separate team may pick, prepare, deliver, clean up, and store them after use.
- The number of recirculating devices controls the amount of WIP in and around the automated system. If there are multiple types recirculating devices for the same operation, only one type should be used for WIP control, with the others kept in plentiful supply.

This pattern is not universal but is found in many different industries. Figure 6.29 shows flows of parts in red and fixtures in blue in a machining example. The automated system consists of five machining centers linked by a shuttle that routes tombstone pallets through one or more of the centers from a raw slab of metal to a machined part. The wisdom of acquiring such a system is not the topic of discussion here.

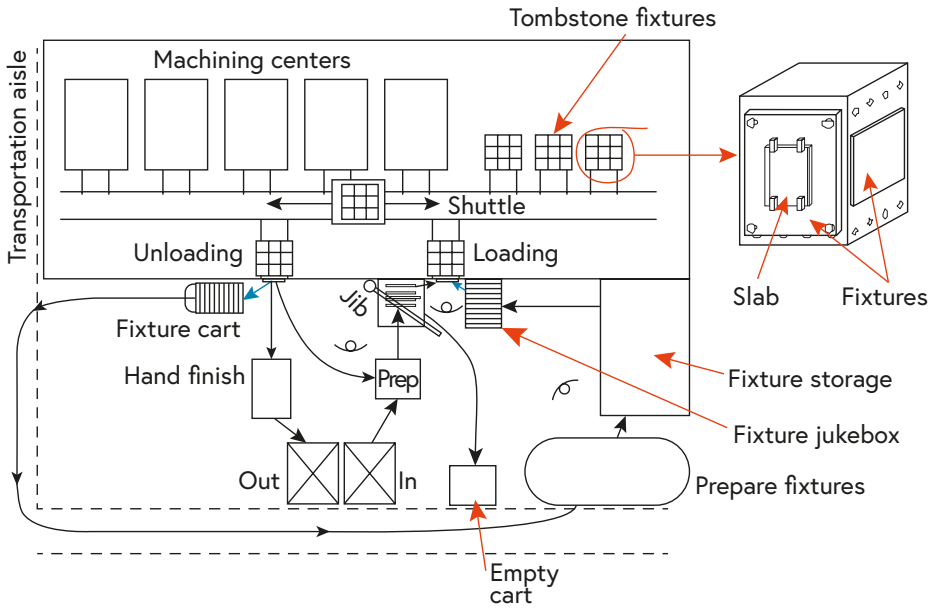


Figure 6.29 Flows of parts and fixtures around automated systems

Indeed, in a green field situation, an analysis that starts from the machining requirements created by a specific demand pattern may conclude that there are cheaper alternatives. What we are looking at is a plant that already has such a system in place, and we are examining how to best manage the flow of parts and fixtures around it. This type of system is typically used where there are many more items than tombstone faces, and fixtures are dismounted after each use to make room for different ones.

In Figure 6.29, the loading station is like an assembly station. The fixtures are presented to the operator in vertical slots, in a sequence that matches that of the slabs, and oriented so that the operator can pull them from the slot and slide them on the tombstone without turning them. They are light enough to be hand-carried through this short distance, because most of the required strength and rigidity is provided by the tombstone.

The raw slabs arrive sequenced and stored vertically in pushcarts. If they are light enough, the operators can pick them up by hand and mount them onto the fixtures. Otherwise, the preparation of the slab consists of drilling and tapping a tie-down hole at the top, which is then used to attach it to a jib crane at the loading station.

At the unloading station, the parts are all light enough to be hand-carried. Those that require machining on the opposite side return to the loading station, while the others go to a hand-finishing area, where they may be knocked out of what remains of the slab on the periphery, straightened, and deburred.

The part flow shown in Figure 6.29 approximates that of a cell without quite achieving it, constrained as it is by the shape of the automatic system, by the need to run parts through more than once, and by the need to recirculate fixtures. The layout is intended to make it possible to operate the area with one, two, or three operators, depending on workload and following the principle that parts must not wait while operators take care of fixtures.

The fixture flow is designed to avoid interference with the part flow. In this case, the amount of WIP in the automatic system is controlled by the number of tombstones, not the number of fixtures. The main issue with fixtures is to have enough to accommodate the variety of items.

6.5.3. **Tool exchange and chip disposal**

In addition to the routine actions described above, there are actions the operators need to take either at fixed times or after a fixed amount of work has been done. These include chip disposal, replacement of worn tools or inserts, and tool presetting. The configuration of most automated systems in the field makes it difficult for the same operator to attend to both these periodic actions and to the ins and out flow of parts and fixtures, as shown in Figure 6.30.

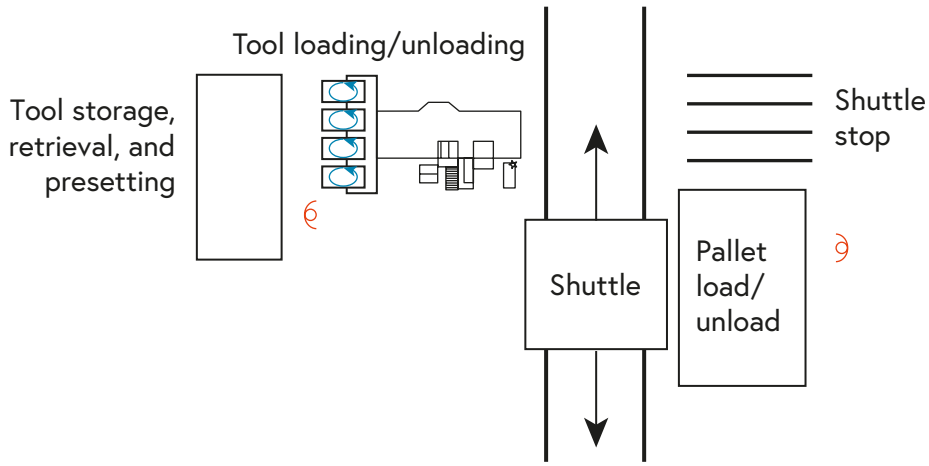


Figure 6.30 Location of tool handling versus part in/out

Figure 6.30 shows a large machine, as can be seen from the size of the operator symbols in red. Pallets are loaded on one side of the shuttle system; tools, on the other. Because of the shuttle, the operators who load do not know which machine will cut the parts. The shuttle track is a practically impassable separation between pallets and machines and the operators on both sides do not see or interact with each other.

Teamwork between both sides is easier if the machine and the loaders are dedicated to one or a small number of products, but it defeats the purpose of having a shuttle that can deliver any pallet to any machine in the system. This is much easier to do with stand-alone machines. They can be set up so that the same operators can do both loading/unloading and tool management. At low volume one operator can do both; at high volume, several operators can rotate among the tasks.

6.5.4. **Tool logistics**

Hardware resources are different in that, if they are not in the machine, the operator has to fetch and load them. One way to eliminate this type of operator intervention is to design the process so that all the resources are permanently loaded.

In the example in Figure 6.29, let us assume that the 5 machining centers have 150 tool pockets each. Then one possible policy is to load the system as follows:

- We assign items for machining by the FMS incrementally.
- We reengineer the process for each item as needed to maximize tool commonality.
- We stop accepting new items when the limit of 150 tools has been reached.

Then we have a common tooling package for all the machines, they are fully interchangeable, and we don't need to worry about loading the right tools for every part. These advantages, however, come at a high price: we give up much of the flexibility of the system. Potentially, we can load 5 times 150 = 750 different tools, and we are only making use of 150. In doing so we are also refusing work that the machine would be capable of if only we relaxed this self-imposed constraint.

The opposite extreme policy is to accept all the work the machine is capable of without regard to tooling issues and to have operators load tools as needed, based on requests from the equipment controller. It does take advantage of the machines' flexibility but increases tooling and labor costs and may result in machining parts on this system that would be better suited for other equipment.

There are also numerous intermediate possibilities. The following is a possible strategy:

1. **Standardize tooling.** Tool standardization is worth pursuing regardless of which type of equipment is used. There is no advantage in using 10 different reamers with overlapping tolerances where one would do, and it complicates tool management.
2. **Sort tools by decreasing frequency of use.** You aggregate the usage of each tool generated by each item, based on planned takt times. Then you sort the list.
3. **Do an ABC categorization of the tools based on the sorted list:**
    A. Common tooling package permanently loaded on all machines.
    B. Tools permanently loaded on one of the machines.
    C. Tools loaded temporarily when needed.

For example, 75 pockets on each machine may be reserved for A-tools, 50 for B-tools, and 25 for C-tools. The system as a whole will then have 75 + 5×50 = 325 "permanent resident" tools supplemented by 5×25 = 125 "guest" tools, for a total package of 450 tools, with only the 28% that are least frequently used requiring operator intervention.

It may also make sense to have an additional category of "super-high-frequency" use tools, with two units permanently loaded, so that the second, redundant one can start cutting when the first one is worn out, and the worn tool can be replaced without stopping the machine.

Summary

This chapter has discussed automation – or more generally, the use of technology in manufacturing. The most essential points covered in this chapter were:

1. Automation is the elimination of human intervention in a process by the means of technology.
2. Automation is a choice and not a panacea.
3. Robots do not play nearly as large a role in automation as appears from news media.
4. The USA principle of automation is a common-sense principle for automation.
5. Toyota's 7 steps for automation is a useful guideline for machining automation projects.
6. Increasing degrees of automation is alienating the operator from the work.
7. Automation solutions should be designed with usability engineering in mind.
8. Do not forget the hidden factory: Tools, recirculating devices, garbage disposal, and so on. All need logistics systems and these too can be automated.

Test yourself

1. When does it make sense to automate a process?
2. What did Elon Musk mean by the tweet "Excessive automation was a mistake"?
3. What is a lights-out-factory?
4. What is a Flexible Manufacturing System?
5. What is a chaku-chaku line? What does it have to do with automation?
6. What is a cobot? How is it different from an industrial robot? When does it make sense to invest in?
7. What is the "USA" principle of automation?
8. What are the 7 steps of automation for machining according to Toyota?
9. What do we mean by alienation in the context of automation?
10. Give five examples of automation in your daily life.

Further reading

Baudin, M. (2007). *Working with machines: The nuts and bolts of lean operations with jidoka*. CRC Press.

Boyer, K. K., Leong, G. K., Ward, P. T., & Krajewski, L. J. (1997). Unlocking the potential of advanced manufacturing technologies. *Journal of Operations Management, 15*(4), 331–347.

Harik, R. & Wuest, T. (2019). *Introduction to advanced manufacturing*. SAE International.

Hounshell, D. (1985). *From the American system to mass production, 1800–1932: The development of manufacturing technology in the United States*. JHU Press.

Jaikumar, R. (1986). Postindustrial manufacturing. *Harvard Business Review, 64*(6), 69–76.

Kagermann, H., Wahlster, W., & Helbig, J. (2013). *Recommendations for implementing the strategic initiative INDUSTRIE 4.0*. Industrie 4.0 Working Group.

Kusiak, A. (1993). *Concurrent engineering: Automation, tools, and techniques*. John Wiley & Sons.

Schwab, K. (2016). *The fourth industrial revolution*. World Economic Forum.

van der Aalst, W. M., Bichler, M., & Heinzl, A. (2018). *Robotic Process Automation*. Springer Fachmedien Wiesbaden GmbH.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

7 Layout

A layout is a floor plan. It may sound simple but do not take it lightly. The layout of a factory is a strategic decision, with a major impact on running costs. It dictates the flow of materials between manufacturing processes, which limits their efficiency. A layout decision is often made for a horizon of 10 years or longer. It is a one-off and very expensive investment. These are some of the many good reasons for planning layouts well.

Manufacturers should design layouts to make production *flow*. Most existing factories are, instead, designed around an architectural facade or power supply and plumbing by outside firms with deeper knowledge of these issues than of flows of materials and movements of people.

Layouts are not just important in factories. They affect our everyday lives too. Next time you are cooking in your kitchen, think about how its layout affects the time needed to prepare a meal. Or, when you are visiting a restaurant buffet, or line up at retail store checkstand, look for bottlenecks and inefficiencies. Layouts also affect the productivity of office jobs, for example in the form of walking distances, double handling, searching, and waiting.

This chapter introduces and discusses the importance and designs of layouts. A factory layout exists at different levels of granularity. A macro layout shows how buildings and departments relate to each other; a micro layout, how production equipment, operator workstations, transport, and open space is organized in a factory. Layouts are planned for current and future product variants, volumes, and variations. Future changes in product mixes, volumes, and technologies can be hard to forecast, but they are key points to consider. Therefore, layout designers need methods and tools to plan and forecast flows. The latest trend is to maintain a digital twin of the factory, with 3D visualizations and possibilities for running simulations.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Explain the differences between types of factory layouts.
2. Select a suitable layout type for a product.
3. Assess the effectiveness of a factory layout.
4. Decide on special layout requirements for a manufacturing factory.
5. Identify monuments in factories and know how to deal with them.
6. Use P-Q analysis to structure shop floor layout decisions.
7. Use layout development tools to assist a layout planning project.

DOI:10.4324/9781351110310-9

7.1. Macro and micro layout

> **Definition**:
> A factory layout illustrates the physical arrangement of departments, workstations, equipment, storage areas, and aisles within a manufacturing firm's facility.

We can separate macro from micro layouts. The macro layout is what you see when you peel off the roof of the plant in the satellite picture and look inside. The micro layout shows how people interact with machines to make materials flow, and is coupled to operator job design. While micro layouts address individual machines or work stations and their relationships inside a production line, the macro layout addresses the relative positions of departments and production lines. When production is spread across multiple buildings and sites, the relationships between these large entities is also part of the macro layout.

Figure 7.1 shows an example of a plant-level macro layout for a factory making moulded components and customized packaging products. It is only intended as an example of a layout, not as an example of a perfect layout.

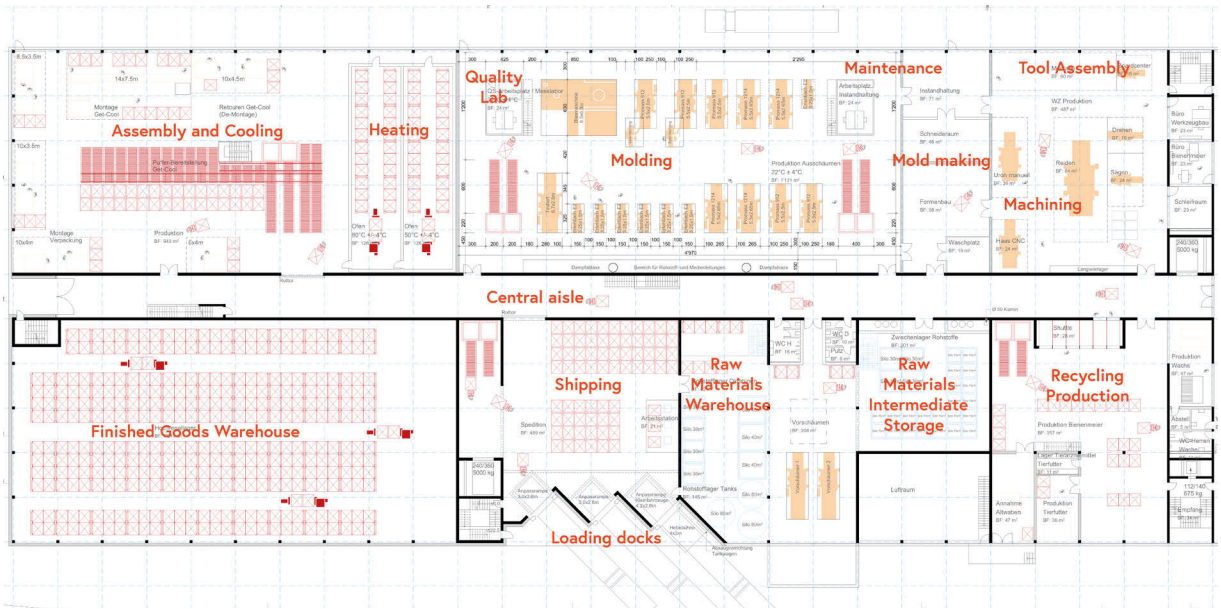


Figure 7.1 Example of a macro layout for a factory (Credit: René Binkert)

The macro layout embeds a configuration of workspaces in the structure of buildings. As such, it is difficult to change and this is why it requires upfront attention. This is perhaps why the bulk of the English-language literature on factory design is dedicated to macro layouts.

The bulk of manufacturing performance, on the other hand, is determined by the micro layout because machines, lines, and workstations are where the processes take place. Inside a factory, the travel distance matters less than multiple handling, accumulating pallets of parts, and using forklifts to move them. If we zoom into the Assembly area in Figure 7.1, we see an example of a micro layout as shown in Figure 7.2.

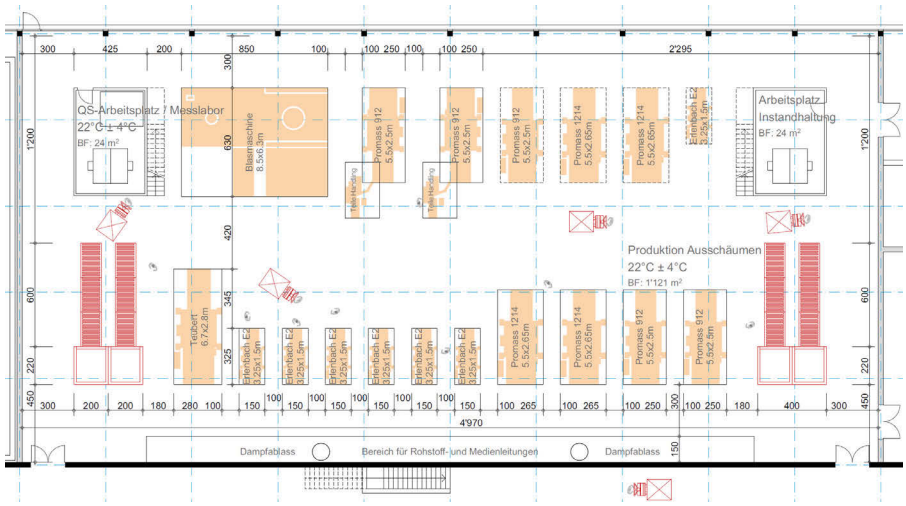


Figure 7.2 Example of a micro layout: Molding area of Figure 7.1 (Credit: René Binkert)

Figure 7.3 illustrates the difference in stakes between macro and micro layouts. In Figure 7.3a, two successive machining operations are located in two distant manufacturing islands. The operator accumulates one pallet of output at Operation 1, and then calls a forklift to move it to Operation 2. Figure 7.3b shows the result of a macro layout improvement bringing the departments responsible for Operations 1 and 2 closer together. This reduces the transportation distance but does not otherwise alter the mode of operation and makes only a limited difference. Figure 7.3c shows a micro-layout change that makes work pieces move *one at a time* from Operation 1 to Operation 2, thus eliminating both the pallet of work-in-process and the transportation step, which is a massive improvement.

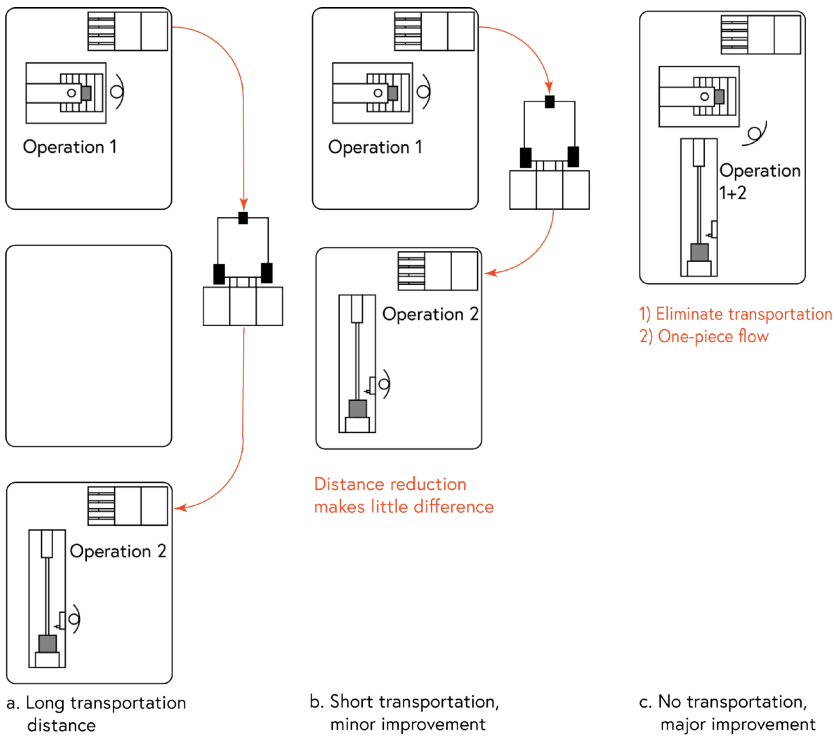


Figure 7.3 Effect of reducing the distance between operations

It is the micro layout that determines how efficiently workers can perform their tasks, how well a production system can be automated, and thus how fast goods can be produced. And micro layout are easier to change than macro layouts. The flexibility needed to respond to changes in demand and new product introductions is easier to build into micro layouts than macro layouts.

Trade-offs made in the micro layout to accommodate the shape of available space left by pre-existing production lines is paid for in longer lead times, additional work-in-process, impaired communications between team members, more transportation between operations that should have been collocated but cannot be, all of which impacts quality through delays in problem detection. Some of these trade-offs are inevitable, particularly when setting up a new line in an existing facility, but a micro layout designed without these trade-offs provides a means of quantifying their effects.

Managers rightly perceive layout changes as expensive interference with running operations. Consider, for example, the factory in Figure 7.4. Therefore, changes in layouts are kept to a minimum. Instead, it is common to locate, for example, new lines where there is available space in the factory. But this "that-looks-about-right" (TLAR) approach is not taking into account the effects of a bad layout on issues like transportation, storage, waiting, etc. More often than not, a detailed economic analysis reveals that just fitting a new line in available space is the wrong priority.



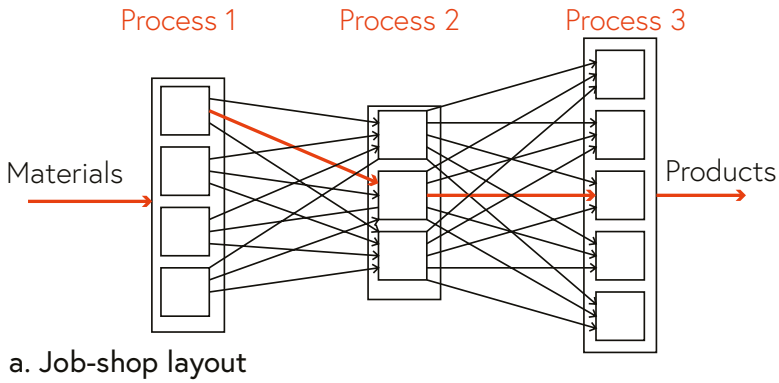
Figure 7.4 Layout changes can be costly projects (Credit: 123RF/yuryrumovsky)

Stop and Think 7.1
Consider the layout in Figure 7.4, how do you think this layout affects daily operations?

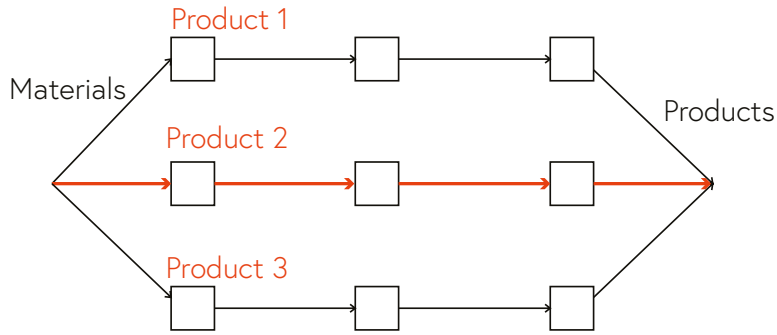
7.2. **Layout types**

Factory layouts can be categorized into three basic types based on a combination of process location and material flows, that often coexist in different shops within a factory (Figure 7.5):

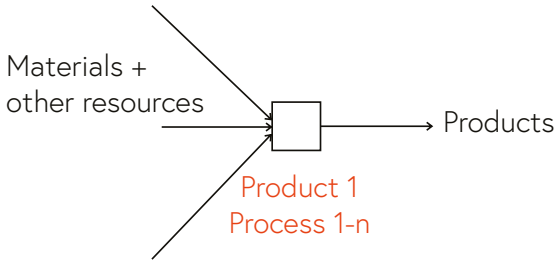
1. **Job-shop layout**, also known as process-oriented layout and functional layout, groups workstations functionally, thus according to the operations they perform. The material can take many routes through the factory depending on the machines they use.
2. **Flow-shop layout**, also known as product-oriented layout, groups different workstations together based on the products they produce. The material flow is dictated.
3. **Fixed-position layout**, also known as stationary layout, clusters all operations around the product. Materials and other resources like machines travel to the product.



a. Job-shop layout



b. Flow-shop layout



c. Fixed-position layout

Figure 7.5 Job-shop layout, flow-shop layout, and fixed-position layout

During production, does the product move or stay in one fixed location? If it's a car assembly line, you usually see it move, continuously. If it's a shop in a shipyard making hull sections for an oil tanker, it stays in a fixed position. Once finished, they roll the hull section over to a dock for assembly with other sections to build up the hull, also in a fixed position. Whether the product is in a fixed position or moves has far-reaching consequences on the delivery of components, fixturing, tooling, skills requirements on operators, and tracking progress.

Job-shop and flow-shop are originally scheduling terms. A *job-shop* is a system in which each job has its own sequence of tasks; a *flow-shop*, one in which all jobs go through the same sequence of tasks. In manufacturing, the terms came to designate types of layouts that support these different modes of operation. A job-shop layout groups machines that are identical or functionally equivalent into work centers, and each job threads its way through multiple work centers. A flow-shop layout groups machines in sequences so that the output of each machine between first and last feeds into the next machine. As a general rule, a job-shop is technically easier to set up than a flow-shop but more difficult to operate.

7.2.1. **Job-shop layout**

A job-shop layout is aggregated by function, make, and model so that, for example, all the Brand-X lathes are side-by-side as are, in another location, all the Brand-Y die casting machines. Each group of functionally interchangeable machines is called a *work center*. The flows between work centers are convoluted and the position of a unit within a process cannot be told from its physical location. Figure 7.6 show examples of job-shop layouts.



Figure 7.6 Examples of job-shop layouts: Machining and sewing job-shops

The unit of work is a *job* that is a set of workpieces going through a series of operations specified and tracked on a physical or electronic *traveler*. In a job-shop, operators are usually specialized in a single operation. A machinist may handle multiple lathes for years and never work on milling machines, which curtails the flexibility and adaptability of the plant. Job-shops are also notorious for burying quality problems in work-in-process (WIP). If 1 out of 10 machines doing the same work is down, the other 9 still work. It makes the job-shop resilient but it also reduces incentives to properly maintain all machines.

Table 7.1 summarizes the pros and cons of job-shops.

Table 7.1 Advantages and disadvantages of job-shop layouts

Advantages	Disadvantages
• Can handle large variations in processing needs • Creates flexible "centres of expertise" • Less vulnerable to changes in product mix and volume • Resilient to machine breakdowns • Simpler utilities: power supply, fluids supplies, exhaust, waste disposal.	• Single-skilled operators • Large buffers of work-in-process • Convoluted materials flows • Complex materials handling • Complex tooling supply • Slow detection of quality problems • Low visibility of products and customers • Low machine availability • Large re-layout costs

7.2.2. Flow-shop layout

In scheduling, a flow-shop is just a system in which all jobs go through the same sequence of operations. As a shop floor layout concept in manufacturing, a flow-shop has consecutive machines and workstations *collocated*. If you identified a fixed route within a job-shop, it might be scheduled as a flow-shop but it would not be a flow-shop *layout*. On a shop floor, flow-shop is a physical, not a virtual, concept.

A flow-shop can take many forms. It can be an assembly cell, a machining cell, an assembly line for large products like tractors, a U-cell for small products, or a continuous flow line making pastries. Figure 7.7 show two examples of job-shop layouts.

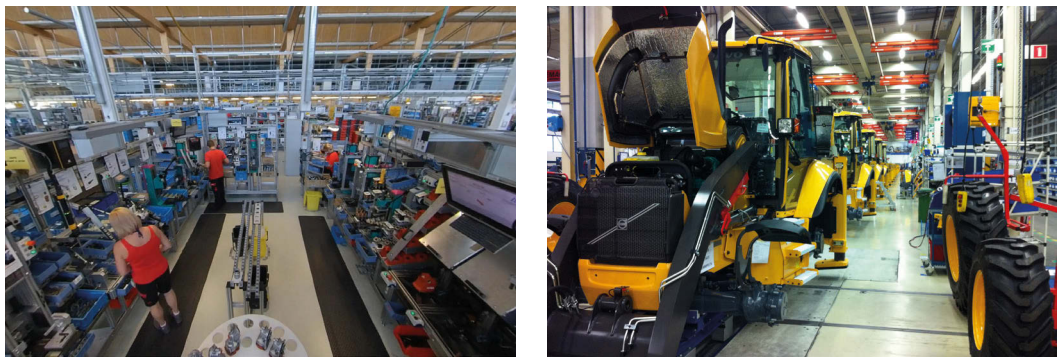


Figure 7.7 Examples of flow-shop layouts: U-formed assembly cell and an I-formed assembly line

There are – as explained elsewhere in this book – differences between these forms of flow-shops, but compared to job-shop layouts, they have some common pros and cons, which are summarized in Table 7.2.

Table 7.2 Advantages and disadvantages of flow-shop layout

Advantages	Disadvantages
• High output-rate • Low unit cost • Specialization of workforce • Reduced materials handling • Efficient utilisation of labour and equipment • Established material flow • Established production planning • Reduced setup time • Reduced buffer levels/work-in-progress • More efficient utilization of labor • Easier to control • Easier to automate	• Low degree of space utilization • Need for extensive education and training of operators leads to increased labor costs • Inefficient for inadequate product families • Increased capital expenditure due to duplication of resources • Vulnerable to changes in product port-folio because of limited flexibility • Requires reorganization for new and different product groups • Can create dull and repetitive work • Individual reward schemes difficult • Vulnerable to break-downs

7.2.3. **Fixed-position layout**

In a fixed-position layout, the product does not move while being worked on. Operators, materials, tooling, work instructions, gauges, and instruments come to the product and move around it to perform operations as needed. Usually, this is used if it is difficult or risky to move the product. This can be the case when the product is too heavy, large, bulky, or fragile, like ships, aircraft, or buildings. It can also apply to prototypes and product made in very low volumes. Figure 7.8 shows two examples.



Figure 7.8 Examples of fixed-postion layout: Heavy machinery and yard operations

Fixed-position layout has also been used with short but variable processes, like desktop computer assembly at Dell. Customers order custom configurations on line. Dell kits each unit in a kitting line and delivers each kit to one of several dozens of assembly stations working in parallel, where one

operator assembles the computer in ~3 minutes. Then it goes for hours of testing. This is unique to Dell, and other computer assembly companies use assembly lines. Fixed-position layout can also been found in bench assembly processes for products like high-end watch production, medical technology, high-technology sensors, research instruments, etc.

The fixed-position concept is simple but delivering all the resources to it is not, and neither is monitoring progress and ensuring that no step is skipped. Table 7.3 summarizes typical pros and cons of fixed-position layout.

Table 7.3 Advantages and disadvantages with fixed-position layout

Advantages	Disadvantages
• Accommodates large products • Low fixed costs • Product must not move	• Requires skilled labor • High variable costs • Complex material flows planning • Complex process routing • Progress tracking challenge

7.3. Role of flow in layout

Flow is key to productivity in manufacturing, but it can be unclear what it means in a manufacturing setting. In manufacturing, we rarely have flow in the literal sense, unless when materials flow in pipes. There is flow inside a continuous casting process but the input is molten metal discrete ladles, and the output is cut into discrete billets. Flow, however, is used *figuratively* for movements of materials, data, and information, people, and money when they have similar characteristics to the flows of fluids.

We don't apply flow, for example, to an individual project, because it ends once a goal is reached, and everything about it is temporary. However, making 1,000 units of a product or product family every day looks more like a flow. If we can organize to do it in a network of operations, each holding discrete units moving at regular intervals, it looks like a flow and we call it a flow, in a figurative sense. It is an approximation, and takt-driven production is the closest this approximation can be.

7.3.1. Flows of materials

Flow is often misunderstood to only apply to assembly lines. Flow and assembly, however, are distinct topics. There are assembly plants that don't have flow, because they do fixed-station assembly or batch-and-queue. Fabrication shops transform workpieces through a sequence of treatments, and many fabrication shops are organized to flow. Historically, the first plants set up to flow were not in assembly but in process industries, flour mills for example. If flow is so natural for assembly, why was the first large-scale assembly line built in 1913? Before then, products like Singer sewing machines were assembled by the millions without assembly lines. Neither type of layout is intrinsically wrong. The question is whether it fits the specific needs of a factory.

A flow-shop layout requires dedicated equipment for a product or product family. By definition, *runners* are products with a sufficiently high and stable demand to justify a dedicated flow line. *Repeaters* are

products that can be grouped into families whose aggregate demand justifies a dedicated flow line even though its members individually don't. The feasibility of such a line, however, depends on the compatibility of the different products. *Strangers* are remaining odds and ends, made in a job-shop.

Whether you can organize production to flow is not a question of size or mix but of *stability*. Plants are not usually laid out and run as flow-shops when, based on the demand pattern, it should have been a job-shop. On the other hand, most traditional plants are laid out and run like job-shops to meet a demand that calls for a flow-shop. Identical machines are in work centers because the architects and engineering firms drawing the layouts have "always done it this way," or because it simplifies drawing and factory planning. They focus on facilities issues, not flows of materials or movements of people. The machines are in neat rows because, 100 years ago, they used to receive power through belts from overhead shafts. By failing to take advantage of flow, these layouts extend lead times, and tank both productivity and quality.

There is always a need for a job-shop work to make strangers. Spare parts for obsolete products are a major group of strangers. The other groups include R&D prototypes, pilot production runs, and sample quantities for OEM customers. Today's stranger may be next year's runner. While strangers *shouldn't* hog management and engineering attention, they shouldn't be ignored either.

Besides the materials that go into products, the layout must also support the removal of scrap generated by the process, and the circulation of tools, fixtures, or dies that wear out or must be changed during setups.

7.3.2. **Flows of information**

Layouts also affect flows of information. In human face-to-face communication, it matters if one can have eye contact and hear each other well. The placement of rooms and machines affects these conditions. Good layouts intrinsically "communicate" with the users through visual cues. For example, layouts can help make it obvious in what direction materials flow, or they can obfuscate it to the extent that even the people working on the shop floor do not know where materials come from and leave to.

Figure 7.9, from tool maker Hilti's factory in Thüringen, Austria, shows an example of offices located right next to the factory floor, separated by glass walls. In this case, the layout helps white collars stay up to date with the status on the shop floor, but also vice versa. It is an example of good architectural design in manufacturing.

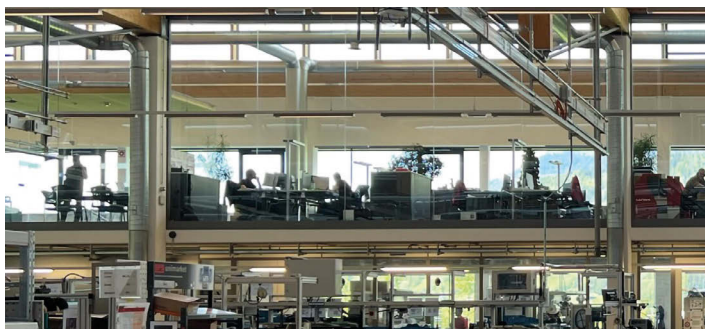


Figure 7.9 Offices located next to the factory floor with glass walls (Credit: Hilti)

7.3.3. **Flows of people**

A factory building is surrounded by grounds that are commonly fenced in and have designated points of entry and exit. Before shift start, lines of employees waiting to flash their badges and cross turnstiles are a common sight. In companies that do not trust their employees, lines waiting to cross metal detectors at shift end are also common. Making them wait to enter or leave the plant is not only requiring them to donate their time but also sends the message that they are not important enough for the company to care about it. More turnstiles may or may not be the best way to eliminate these waits. Turnstiles are effective at tracking people in and out but slow. Portals that can read badges as employees walk in might be a possible alternative. The role of plant security is usually focused on preventing unauthorized entries but it should also facilitate authorized entries.

Once inside the factory gate, employees often find themselves sharing space with trucks or forklifts, but people and trucks do not belong in the same environment. Not only do they interfere with each other but they are accidents waiting to happen. This is easiest to avoid when designing the plant but an existing plant can sometimes be fixed with minimal construction such as clear marking of aisles and traffic routes.

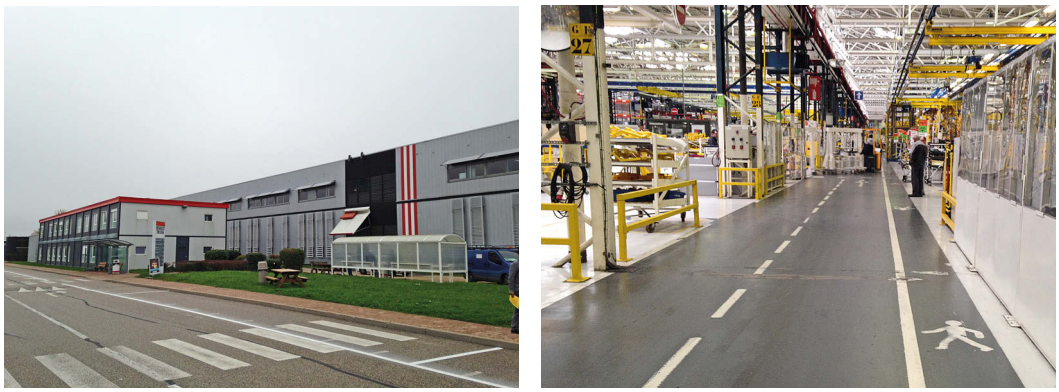


Figure 7.10 Separating flows of people and vehicles outside and inside the same factory

The flows of people are also very much dictated by locations of rest rooms, indoor and outdoor break areas, wardrobes, and cafeterias. A factory designer should not guess where most people will work. It is information to share during planning.

Factories also have to accommodate visitors, who may be government inspectors, customers, suppliers, professional societies, or schoolchildren on field trips. This requires a visitor center, trained guides, visitor paths on the shop floor, and marking for areas that are off limits. Box 7.1 shows the example of Porsche's plant in Leipzig, Germany, where the layout design is carefully planned from the perspective of customers and visitors.

Box 7.1 **Factory tour in Porsche Leipzig**

While guided tours are a secondary activity in almost every factory, there are exceptions like Porsche Leipzig. As seen in Figure 7.11, the visitor center is the highest structure and the shop floor looks like a showroom.

For a fee, buyers are allowed to pick up their cars at the factory. They go on a guided tour of the assembly floor, spend an hour on the test track with a professional driver briefing them on taking care of their new Porsche, buy souvenirs in the shop and lunch on the top floor of the visitor center. None of the dirty processes like stamping or painting are visible to visitors. The final assembly area looks like a showroom. For hosting suppliers, there are often meeting facilities attached to the front lobby and accessible without going through security.



Figure 7.11 Porsche visitor center and factory in Leipzig, Germany (Credit: Porsche)

When factories are built as combined showrooms and manufacturing facilities, there are special requirements to layout. Design, tidiness, ambience, and cleanliness often take center stage. It is, however, also important to ensure that tours do not disturb work flows, that employees' privacy is respected, and that industrial espionage is hindered. When these things are catered for, few sales offices can create the same levels as customer loyalty as the factory. When customers see the product being made with craftsmanship and passion they buy an experience, not just a product.

7.3.4. **Flows of vehicles**

Factories also receives flows of materials from outside. Trucks interact with the plant through docks. Where truck heights are standardized, docks are usually built to match the height of the truck beds, allowing content on casters to be wheeled in an out or forklifts to pick up pallets inside the trucks. Most trucks are loaded and unloaded from the back and therefore park with the back facing the dock. The downside of this method is that it is last-in-first-out (LIFO). Side loading trucks offer random access to their loads and can be loaded and unloaded faster, but require longer docks that extend inside the building if protection against the weather is needed.

Some plants have separate rows of docks for unloading and loading, on opposite sides of the building. Materials received on one end flow through the building and ship as finished goods at the other end. While this separation has its advantages, when the same trucks are used for receiving and shipping, they need to be moved around the building after unloading. This is also problematic when a shipment commingles products made inside the plant and outsourced, because the outsourced products are unloaded on the receiving side and traverse the entire plant without undergoing any transformation to be loaded at the shipping. This unnecessary transit is eliminated if unloading and loading are possible on the same side.

The means of transportation used by employees to commute to work also have an impact. Where they use public transportation or company shuttles, the factory buildings can occupy 80% of the grounds; where they use their own cars, the parking lots may often be larger than the factory. Some factories have gaps between shifts that provide slack to catch up when production is late and to maintain the equipment. Others like to have overlapping shifts for personal communication between the outgoing and incoming teams. If employees use their own cars, then the parking lot must be large enough for two shifts, as in Figure 7.12.



Figure 7.12 Toyota Motor Manufacturing, United Kingdom in Burnaston (Credit: Toyota)

7.3.5. What can we learn from Google Earth?

The interactions between a factory and the rest of the world are expressed in the shape of its buildings, its position with respect to transportation resources, the arrival and departure of people, receipts of materials, and shipments of finished goods. Google Earth images of factories reveal many of these features such as external logistics of workers and materials, factory size, and number and type of buildings.

The examples in Box 7.2 and Box 7.3 highlight examples of effective designs and of costly mistakes that are sometimes impossible to fix once the plant is built. The consequences of neglecting them in plant design include accident-prone areas shared by people and trucks, traffic jams of trucks on a too narrow access road, and employees waiting in line to check in or out of the facility.

Box 7.2 Comparing Boeing versus Irkut with Google Earth

The satellite photos in Figure 7.13 contrast an American and a Russian plant both involved in aircraft manufacturing. In aircraft plants, the finished products fly off a runway. The two plants are comparable, because Renton builds the Boeing 737 and Irkut produces the MC-21, which is intended to compete with the 737. From space, Boeing Renton looks clean and modern; Irkut's plant, old and decrepit, but it's a Cold War-era plant, and its appearance to satellites flying over is intended to camouflage rather than impress.



Figure 7.13 Boeing Renton (left) and Irkut (right) (credit: Google Earth)

Stop and Think 7.2
Go to Google Earth and look up a local factory you know of and compare it to the factories in Box 7.2 and Box 7.3. What are the most striking differences? What can explain it?

Box 7.3 **Comparing Dell Austin versus Intel Rio Rancho with Google Earth**

The Dell facility in Austin, TX, (Figure 7.14, left) was comprised of 3, identical leased buildings, each surrounded by dozens of trucks, and the entire complex was laid out so that trucks and people would never cross paths. It assembled tens of thousands of computers everyday, receiving large volumes of packaged electronic components and shipping equally bulky desktops and servers. It is now closed.



Figure 7.14 Dell, Austin, TX, in 2012 (left) and 2021 (middle) and Intel, Rio Rancho, NM (right) (Credit: Google Earth)

Dell chose to lease buildings for which alternative tenants could easily be found, which made it easy for them to pull out. The population of the parking lots reveal that the three buildings were put to different uses. Building 2, with the full lot, assembled 20,000 desktop computers/day. Building 1 assembled a few hundred servers per day. Building 3, with the nearly empty parking lot, was a warehouse serving the other two buildings. The middle photo in Figure 7.14 shows what became of the site as Dell actually pulled out.

The Intel Rio Rancho plant is built on land owned by Intel that had been gradually filled up with production facilities over more than 30 years. At its peak, this was the largest semiconductor fabrication plant in the world. Its latest fabrication area, Fab11X was the state of the art in 2011, but no longer is, and the site is now on the decline. Almost no truck is visible in the picture of the Intel plant. That's because transforming raw silicon wafers into chips typically requires 200–400 process steps that does not consume or produce a great bulk of materials.

7.4. Layout planning and design

Because layouts strongly affect how we can operate the factory, layout planning and design is an essential skill. Many aspects must be carefully considered – from flows of materials, information, and people to security and environmental footprint. Besides these technical aspects, success is dependent on good project management and consistent organization in all phases of analysis, conceptual design, detail design, realization, and commissioning. The process begins with the original need and extends through several planning stages as explained in the following sections.

7.4.1. Design the layout from the inside out

The main goal of the factory layout is to ensure an efficient flow of materials, information, and people through the whole production system. Therefore, the first principle in any layout project is to plan *from the inside out*. The physical properties of the production and logistics system – people, product, process– should play the central role. Layouts must consider different factors including profitability, future changes, and the needs of the workforce.

To make the plant profitable, the layout must support effective and efficient manufacturing. This translates to multiple objectives:

1. **Quality.** As discussed above, quality problems tend to be buried in WIP in job-shops and exposed in flow-shops. The layout designers need to think through the specifics.
2. **Short lead times.** The layout should enable smooth product flows through the process, with few or no points of WIP accumulation. This affects the responsiveness of the process and the speed with which problems are detected – and therefore the length of the organization's cycle of learning.
3. **Operator engagement.** When operator job design is an afterthought, operators are underutilized, discouraged from acquiring new skills, and prevented from working as a team in routine operations and improvement.
4. **Resilience.** Operations are routinely disrupted by changes in volume, operator absences, machine failures, departures of experienced operators, or arrivals of new ones, engineering changes and new product introductions. The layout must support responses to these disruptions. A possible way to ensure flexibility is to provide enough space for potential rearrangements and for increasing or reducing capacity with more or fewer operators and machines.
5. **Few transportation steps.** Inside a plant, the number of transportation handovers matters more than their lengths.
6. **Appropriate handling.** Supplying materials and components to production lines requires choices of vehicles, handling methods and containers. It also requires a transportation network inside the plant.
7. **Effective use of space.** Using space effectively rarely means filling up the available space. Layout designers often need to concentrate production in a fraction of it to leave room for expansion.
8. **Reconfigurability.** The business environments of most manufacturing companies is turbulent, and layouts must consequently be reconfigurable.

A factory layout must respect human needs, so that workers can operate the production system in a safe, efficient, and effective way. It therefore needs to make sure that working conditions do not put the health of any person at risk, be it an employee, a visitor or other people in contact with the factory. Particular attention has to be paid to health and safety standards as well as to ergonomics.

An ergonomic workplace takes manufacturing operations, the arrangement of objects, and anthropometric factors into consideration in order to protect the worker from stress or strain. Further, the layout has to make sure that communication between employees is facilitated. Workstations that require much communication with other processes should be centrally positioned and grant access to all involved employees.

The layout needs to support the work of all involved with the factory, whether they are employees, contractors, vendors, customers, or government inspectors. Shop floor operators need the following:

- Access control methods to the facility that do not cause waiting lines at shift start and end.
- Access paths to the building that are not shared with trucks or other vehicles.
- Lockers for personal items like purses, in a change room if special work clothes are required.
- "Green corners" for breaks on the shop floor.
- Lunch rooms close enough that they don't have to spend half their breaks going back and forth.
- Enough rest rooms to avoid waiting lines.

Attention to these details is not only a means of ensuring that operators can focus on their work. It also shows management concern for their well-being in concrete ways that motivate them to stay with the company and engage in making it successful.

Not all visitors are usually allowed to see the entire plant or walk its floor and, as a result, meeting rooms and visiting paths for outsiders are part of the layout. Figure 7.15 shows the visitors' gallery in the Siemens factory in Amberg, Germany, during a visit of Angela Merkel.



Figure 7.15 Chancellor Angela Merkel visiting the Siemens Amberg factory in 2015 (Credit: Siemens)

7.4.2. **Usual errors in layout projects**

In many layout projects that concerns the placement of new machines or lines in a factory, the first words out of the project manager's mouth are: "This is the space we have to work with," and he rolls out a facilities blueprint. "Management won't spend for new brick and mortar, so we have to make do with this building!"

Behind this action are both a philosophy of how lines should be designed and a particular model of the factory, embedded in the blueprint. It shows the footprint of equipment in the rest of the plant, the locations of sustaining columns, and a patchwork of blank spaces, as shown in Figure 7.16.

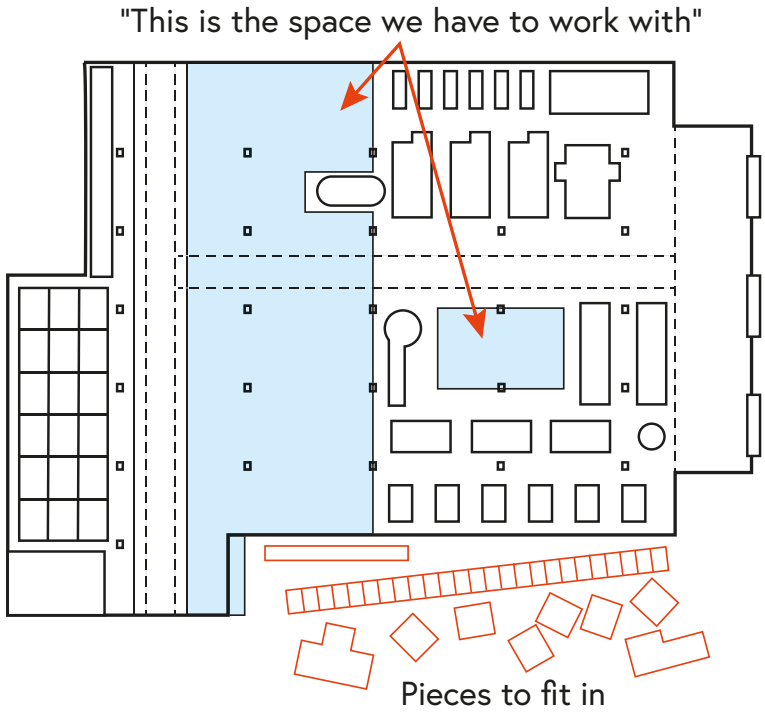


Figure 7.16 Typical management error in layout projects

It also comes with an injunction from the Facilities department to provide the final layout within two days so they can start ordering equipment and services, and install power outlets, air pipes, and drains.

The shop floor design process is then reduced to finding a way to locate all the necessary equipment with the required clearances in a space that is oddly shaped and often discontinuous. The blueprint is the perfect graphic representation for such an effort:

1. It shows the available space.
2. It shows nothing else about the process.
3. It is self-explanatory to anyone with basic map reading skills.
4. It is easy to generate. Facilities prints the blank layout and the equipment shapes, and the team can generate the layout with scissors and glue.

The typical result of such an effort is shown in Figure 7.17.



Figure 7.17 Example of a line design based on the "available space" principle

This design meets the objectives that have been given to the team: "Fit it all within the space available and be quick." From the broader perspective of how well this line will serve to produce 60% of the company's revenues over the next two years, however, it may not look quite as appealing, as can be seen by following a product unit from beginning to end, an exercise every layout team should do. The result is the "spaghetti map" of Figure 7.18.

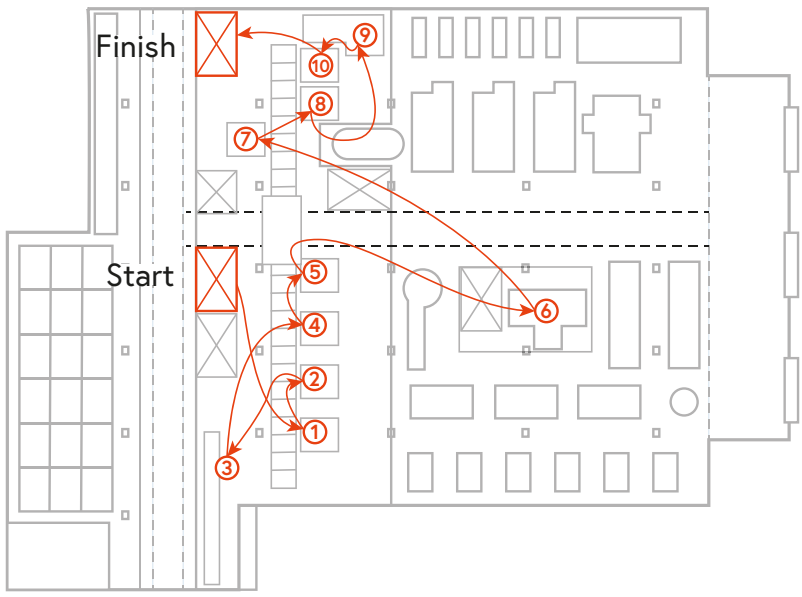


Figure 7.18 Spaghetti map of the assembly process

While Figure 7.18 reveals a nonsensical flow pattern, it is a mild case compared to the inextricable webs found in actual factories. In fact, these "spaghetti maps" serve one purpose only: To show management how convoluted the flow is. It is difficult to draw accurately, and should only be done on a sample basis, to achieve the desired result. There is no need to draw spaghetti maps of every process everywhere in the plant.

In this type of layout, operator jobs are an afterthought. The designers simply make sure that all machines and assembly stations are accessible, but do not have the time to worry about the details of the work operators will do, how they will move and how parts or tools will be delivered to them. As a result, they end up using more operators than needed, some operators are busier than others, and the number of operators cannot easily be adjusted up or down as demand changes.

Having this type of flow pattern for all products throughout the plant has consequences:

- Work-in-process (WIP) inventory buffers are needed every place the product units do not flow to a contiguous station, and a transportation service is required.
- Visibility is destroyed as the paths followed by different products are intertwined.
- Separation prevents operators working as a team and communicating, particularly about quality problems.
- The narrow focus on floor space actually *inflates* the space requirements by 30% to 40%.

Saying that floor space should not be the design driver is not saying that it should be ignored. The correct approach is for the team to first design the best line it possibly can, and then look for a place to put it, which may involve the improvement or relocation of other lines, or even new brick and mortar if economically warranted.

It often involves trade-offs, but at least the team knows what it is giving up and can be selective about it. The convenience of the Facilities department is also a secondary concern. To facilities engineers, the work of designing a line looks like child play, compared with the "real" work they do. In reality, however, they are there to support production, and should not be empowered to rush the line design process. If Facilities has to undo and redo some part of the line installation work, it is a one-time cost, but if the line is badly designed, it is a daily recurring cost.

> **Stop and Think 7.3**
> Why is it more important to consider *future* flows of materials, people, and information than the current facility blueprint during a layout project?

7.4.3. Layout planning process

The planning of a factory layout is a complex project aiming at installing a new site or an internal change or expansion of an existing manufacturing site. It requires the composition of an interdisciplinary team with an ideally sound background in factory planning and additional expertise in assembly and manufacturing technologies, process planning and control, material flow and logistics, business administration, construction design and engineering, and infrastructural planning. However, in some special cases, experts from further fields such as environment, meteorology, city planning, and so on, still need to be consulted.

As shown in Figure 7.19, the layout planning process can be divided into three main phases before the layout can finally be installed.

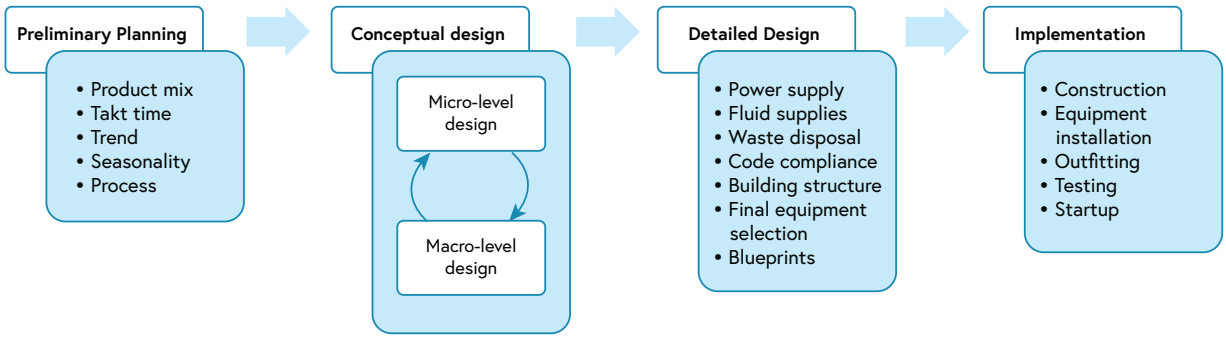


Figure 7.19 Layout planning outline

Step 1. Preliminary planning

Most layout design projects are within an existing facility, referred to as a "brownfield," as opposed to a factory yet to be built, or "greenfield." For greenfield, the choice of a location is a strategic decision, based on objectives ranging from the pursuit of cheap labor to that of proximity to a market, and influenced by factors like the existence of industrial infrastructure in a location and incentives offered by local governments. It is a major topic but outside the scope of this chapter.

Whether in a greenfield or a brownfield, the initial conditions to start designing a layout include knowing the product mix, takt times, trends, and seasonality of the demand, as discussed in Chapter 3. Engineering teams are sometimes pressured to generate layouts without this basic information, generating problems that are more difficult to solve later than upfront.

The process, at least in terms of the physics and chemistry to be applied to the materials, needs to be worked out before embarking on layout design. The issues of process design are addressed in Chapter 4 and Chapter 5. There are exceptions where speed matters so much that a manufacturer is justified in taking the risk of designing and building a production line *before* the process is ready. With government help, Pfizer-BioNTech and Moderna took this risk in 2020 with their COVID-19 vaccines and it paid off in starting massive vaccinations earlier than anyone had predicted.

Step 2. Conceptual design

The conceptual layout is focussed on the flows of materials and the movements of people, including both the micro and macro level. The micro level determines most of the line's future performance and therefore deserves the most attention but neither can be completed without input from the other. For example, separating departments for diecasting and assembly is a macro-layout decision, but it is driven by micro-layout issues, like the need in diecasting to handle molten metals, hot dies, and hot workpieces, with the special training and personal protective equipment that it requires and that assembly work doesn't. Conversely, in a brownfield situation, there is a pre-existing *macro* layout.

During conceptual design, the first phase is to develop the best micro layout the team can envision for flows of materials and movements of people, disregarding any macro-layout issues. It focuses only on the necessary functional elements of the production system – machines, tooling, operators, instructions and programs. It must respect the laws of physics but is not constrained by any existing or planned infrastructure.

Once you have this design, you confront it with the existing or planned infrastructure in the macro-level design. It then serves as a basis to quantify any trade-offs that may be needed when building the actual line. Companies that operate similar type of factories across the world, can benefit from developing and maintaining a global standard that can be adapted to different needs. An example is Toyota's Global Body Line in Box 7.4.

Box 7.4 **Toyota's Global Body Line (GBL)**

Toyota developed its current approach to the welding of car bodies – the Global Body Line (GBL) – in the 1990s and started deploying it in 2006. With assembly plants distributed worldwide working at vastly different takt times, they needed a line design to execute the same process with robotic welding where justified, manual welding where not, and transitions from one mode to the other as conditions changed.

Of the many challenges of the GBL, the most difficult one was fixturing. The parts needed to be positioned accurately enough for robotic welding, yet provide enough operator access for manual welding. Over the years, across locations and products, and generations of welding robots, the most enduring part of the system is the fixturing concept, not the actual machines used.

Based on the process design, relevant input variables for the initial layout are:

- Type, number, and needed space of the functional units within the production system, such as machines, molds, dies, fixtures or jigs, and operator workstations
- Relations between the functional units
- Qualitative and quantitative interdependencies of the functional units such as functional schemes, material flow diagrams, closeness rating analysis, transport matrices, etc.
- Basic structural shape of the layout type

A key point is that the selection of machines down to the level of make and model, is based on the process design and the required takt time. It must be made as late as possible and just early enough for the machines to be delivered, installed, and tested prior to production start.

To figure out the space required, you first need an idea of the layout and equipment that is needed. It's not something you can simply calculate. Based on an equipment set, you can calculate lower and upper bounds, but you can't have a precise value until you have a layout. In addition to its own footprint, each machine needs space for work pieces coming in and out, for operators to attend to it, and for maintenance technicians to service it.

For each machine, this space can be identified on the blueprint based on the machine's specs, workpiece dimensions, and local regulations. But this space can be shared among machines, so that, if we add the total space required for all machines in standalone mode, we get an upper bound of the space required. Then, we try to locate the machines so that the sharable space is actually shared among as many machines as possible, to get as close as possible to the lower bound in space requirements, while pursuing other dimensions of performance.

At conceptual design, drawings were for communications between production engineers, managers, and production operators, and are not intended to drive construction work. They can be drawn by hand on white boards or flip charts, they can be simulated in miniature in LEGO blocks or full size in cardboard, or they can be drawn with 2D or 3D charting software the engineers are used to. Whichever method is used, drawings or models should be generated, modified, and saved as needed by team members.

Step 3. Detailed design

The *real layout* planning starts with an adjustment of the ideal layout based on constraints. It involves confronting the ideal design with a blueprint of the existing or planned facility within which management expects production to take place. Machines or work stations may have to be pulled apart to avoid building columns, and the flow may have to take a 90 or 180 degree turn, and breakpoints may be introduced in the flow to support the use of shared equipment. These changes may degrade the performance of the line, in ways that can be anticipated and quantified. Where these performance losses are sufficiently high, the organization may choose instead to modify the facility so as to remain closer to the ideal design.

First, you create various different real layout alternatives. Second, you include the logistical elements needed to link the functional units to guarantee a material flow within the structure of all available surfaces and spaces. These two steps are recommended to be partly run simultaneously, as it will allow the planning team to consider potentially needed space for logistical elements.

Compared to the ideal layout, the real layout is more detailed and represents a "realizable" design solution. It depicts a realistic spatial arrangement of the functional units considering all underlying restrictions:

- **Manufacturing flows.** Flow of materials, personnel, information, and energy, also allowing for all supply and disposal processes.
- **Manufacturing process.** Process structure, manufacturing equipment, manufacturing process-related conditions, organizational structures, work design.
- **Spatial structures.** Site structure, facility structure, constructional technology, surface utilization.
- **Logistical principles.** Arrangements according to the logistical requirements.
- **Communications**. Communication enabling or facilitating structures.

The goal of the detail layout planning phase is to iteratively modify and detail the real layout so that the design is finally ready to be implemented. Thereby, the exact location of each machine, workstation, aisle and other logistical area within the layout is defined. While defining the detail layout, the efficient and flawless interaction of humans, equipment and the material has major priority. The following aspects therefore need to be considered:

- Positioning of the human in the work and manufacturing process
- Allocation of the workpieces
- Lightning, climate, and noise control

- Access to utilities
- Space to allow material flows and storage
- Industrial safety, fire prevention, ESD, and explosion control
- Maintenance, repair and operations facilitating arrangement of the equipment
- Meeting environmental/ecological requirements
- Meeting legal requirements
- Integration of informational and communicational systems

Depending on the individual characteristics of each project, the scope of work can vary strongly. In any case, the detailed layout planning results in a huge variety of different documents. Among others, these include the final layout, which can be provided in a form of a series of drawings, a CAD drawing, or a three-dimensional model of the layout including the machines, equipment, and other physical features of the layout. Sometimes a complete digital copy of the factory is built, allowing simulation of flows and scenarios. Further documents can include explanations, functional descriptions, task descriptions, calculations, requirements lists, installation overviews, etc.

At this stage, the design drawings are intended to drive construction, and need the corresponding details and geometric accuracy. Computer Aided Design (CAD) software is called for. An example of a CAD drawing is given in Figure 7.20. In the software, it allows for seamless drill-down in layers of the layout. Given its emphasis, this work is commonly contracted out to architectural and engineering (A&E) firms. The team from the manufacturing company evaluates the detailed layout variants and makes the final selection.

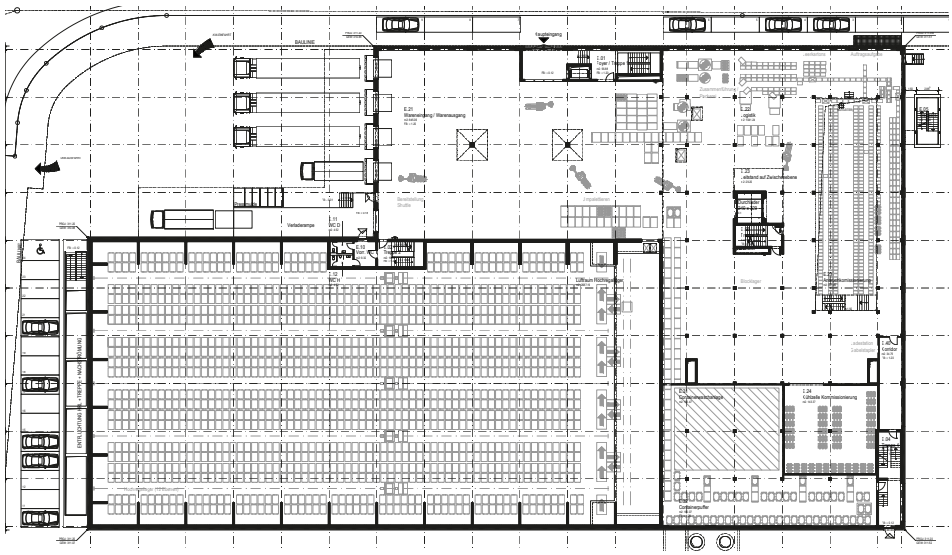


Figure 7.20 A macro layout suggestion of the first floor of a factory for pharmaceuticals (Credit: René Binkert)

Step 4. Implementation

Regardless of how much of the work it contracts out, the manufacturing company's team manages the project and its leader is responsible for its success. Depending on scale, the implementation of a factory building project can take days with a dozen people or years with hundreds or thousands involved, and the planning and organization varies accordingly.

Rather than using project templates or looking forward in time from the present state, the safest way to avoid omissions is to start from the end result, as expressed in the detailed design, break it down into elements, and then break down these elements into lower-level elements, etc., down to the simplest tasks. Precedence constraints and resource requirements fall out of this analysis along the way, supporting the development of installation instructions and timelines.

After construction, installation, and testing of a new facility, the employees who will work with or in it undergo training before the production is launched. Production starts on a new line with a ramp-up period, during which the engineering team stays involved in debugging the process.

7.4.4. Special requirements

Although factory engineers would prefer fully flexible factories, where capacity can be changed easily and every machine can be moved and configured according to actual demand, there are a number of special requirements in manufacturing that constrain layouts, such as the following:

1. Product characteristics
2. Foundations
3. Utilities access and distribution
4. Fire protection
5. Cleanrooms
6. Electrostatic discharge (ESD)
7. Explosion-proof areas

Product characteristics

The following product characteristics are particularly important for the layout planning:

- **Product dimensions**: Large products like airplane fuselages, truck chassis, or elements for buildings require large spaces for manufacturing. Small products such as watches, door locks, or microchips create other challenges, for example technology to handle tiny parts.
- **Product volumes**: External and internal logistics must be designed to support the daily numbers of incoming and outgoing trucks.
- **Product weight**: Heavy products may be handled with devices like bridge cranes that can become bottlenecks. This can sometimes be avoided by using alternative means of moving the items.
- **Product shape**: Fluids require containers or pipes to move. Piping is operationally simpler, with less handling of hazardous materials. Pipes, on the other hand, are more difficult to reconfigure and to clean during product changeovers.
- **Product value**: The product value describes how precious each unit or mass of the product is. If your raw material is a kilogram of lead, you would plan your capacity and layout differently than if it is a kilogram of gold.
- **Product perishability**: You do not want to hold large inventories of perishable products like medicines, meat, or dairy products. You also need special storage facilities and devices for cold storage and transportation.
- **Product features**: Explosive, corrosive, frozen, fragile, sharp, or poisonous materials need special handling. Often, they must be stored and treated separately from other materials by specially trained operators.
- **Product life cycle**: If the product life cycle is short, such as in mobile phones and fashion, you need flexible layouts.

- **Product complexity**: Very complex products can require highly skilled labor. The accessibility and cost of skilled labor is an important factor in locating factories.

Considering the list above, and the elements that follow, the first thing we should understand is that there is no "optimal" layout. It doesn't exist. One can try to "optimize" on a certain parameter, such as material distance travelled, but – when it works – it is suboptimization, not optimization, and it is never valid for long. There are many good reasons to be skeptical when you hear "layout optimization."

Foundations

Many machines are either too heavy or too sensitive to vibration to be placed on normal industrial floors. These machines require additional foundations. Because special foundations are much more expensive than regular floors, they are often constructed exactly where the machines are planned to be. This can create islands of reinforced foundations around the factory and create logistical challenges for future expansions and machine relocations.

Utilities

Factories in general requires access to electricity, gas, and water, and other fluids. Some machines and processes require more than others. For example, in a steel smelting plant, the oven consumes most of the energy. In composites, the autoclave requires the most electricity. Machines that generate heat may need more access to coolants. A paint shops requires more access to gas, electricity and water than an assembly cell. All these requirements must be taken into account when planning and designing the factory layout. The results will usually be some kind of *monument*; an inflexible process that cannot be moved without starting a complete reconstruction of the factory.



Figure 7.21 Semiconductor fabrication requires special fundaments for sensitive processes, special utilities for air and energy supply, an ESD-safe environment, and cleanrooms (Credit: Bosch AG)

Fire protection

Although legislation varies between countries, all industrial buildings are normally required by law to be protected against fire. Besides the usual fire protection systems such as installed sprinklers and alarms and available fire extinguishers, factories also have *structural fire resistance* such as firewalls and ceilings. Large factories are cut into rooms separated by firewalls able to contain a fire for a set duration, for example, 60 minutes. This has several implications for layout planning. For example, if one department is separated from its supplying department by a firewall there will be material transportation moving through defined doors if transported with forklifts, AGVs, or people, or hatchways if on conveyors. Doors will define logistics patterns, create noise, safety risk areas, and congestion points. The walls will also impede communication and teamwork between operators on both sides.

Cleanrooms

Factories that make precision products or products that are consumed by humans and animals need cleanrooms. A cleanroom is a controlled environment that has a low level of pollutants, such as aerosol particles, dust, airborne microbes, and chemical vapors. Cleanrooms can be found in a wide variety of manufacturing industries, including medical devices, pharmaceuticals, biotech, semiconductor manufacturing, fine-electronics industry, jewelry and watch making, military equipment, aerospace, optics, and food.

A cleanroom has a controlled level of contamination that is specified by the number of particles per cubic meter at a specified particle size. In addition to strict cleaning standards and routines, cleanrooms are specially designed to reduce the risk of contamination and pollution entering the rooms. Cleanrooms have entrances that seal properly when closed, controlled temperatures and humidity, controlled ventilation, and constant overpressure. Air entering the cleanroom flows through high efficiency particulate air (HEPA) filters or ultra low particulate air (ULPA) filters.

Factories with cleanrooms also need to build wardrobes, washing facilities, and gateways to enter and exit the facility. Employees need to wear cleanroom clothing and logistic systems must be designed to supply cleanroom equipment and dispose potentially contaminated clothing and material. The constraints are similar when the objective is to protect operators from the products instead of the products from operators, when working with hazardous chemicals or radioactive materials.

> **Stop and Think 7.4**
> Can you think of any way to reduce the need for cleanrooms without lowering the standards for the end product?

Electrostatic discharge

Factories involving assembly work on electronic devices are at risk of electrostatic discharge (ESD). ESD is a sudden flow of electricity between two differently charged objects. It can be caused by contact, a dielectric breakdown, or an electrical short. Lightning is an example of ESD. ESD protected areas (EPA) should be properly marked (Figure 7.22). EPA areas can be limited to a workstation or

encompass a whole building. To avoid damage to people and products, assemblers touching the products wear grounded conducting wristbands, walk on grounded anti-static mats, or wear ESD shoes. Humidity reduces the risk of ESD. Also materials going into the EPA and the equipment in the area must be cleared for ESD risk. Materials can be packed with ESD-safe packaging material. There are international standards for EPA defined by the American National Standards Institute (ANSI) and the International Electrotechnical Commission (IEC).



Figure 7.22 ESD warning signs

Explosion proof areas

Some materials are explosive, whether by intent or not. Industries working with a range of powder materials are at risk of explosion. Although some powders are at higher risk of explosion than others (e.g., aluminum, magnesium or titanium powders), all powders pose a risk when fine-grained enough. While often thought of as harmless, grain, flour, sugar, coffee, and cocoa can also create dust explosions. When particles are mixed and dispersed with oxygen in the air, dust explosions can ignite. Over the years, bakeries have burned down and people lost their lives as the consequence of dust explosions.

Any area at risk for explosion needs protection against it. The usual industrial marking for explosion proof is "Ex." In an ex-proof area, more than yellow marking on the floor is needed. Any equipment going into the area has to be ex-proof. This can easily increase the cost of the equipment by a factor of 5 to 50. Consider Figure 7.23: The two shop floor clocks appears similar, but the one to the left is ex-proof (see label). The cost of the clock to the right is USD 20, the one to the left costs USD 800! This was a simple example with clocks, but imagine if you want to fully automate an area that is explosion proof versus one that is not.

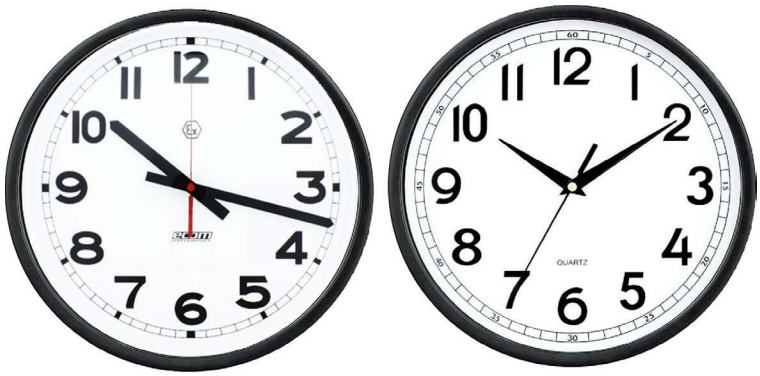


Figure 7.23 Two apparently similar but very different shop floor wall clocks

Note that dust explosions are particularly relevant to additive manufacturing where raw materials are powder-based (e.g., "metal printing"). Hence, knowledge about explosion-proof areas will become more important and relevant when additive manufacturing takes over more tasks from conventional manufacturing processes.



Figure 7.24 An environment as "innocent" as confectionery production can be dangerous when fine-grained powders mix with air and can create dust bombs

Operations with a risk of explosion are often located in open spaces outside the main building, or on its side of the building rather than the center where explosions would hurt the most people and do the most damage.

7.4.5. **Design for safety**

When a workplace is safe, the workforce takes it for granted. It is only when accidents happen or employees compare it with the workplaces of less safety-conscious employers that they take notice. Safety by itself is not a motivator but unsafe working conditions are a demotivator.

Safety is often treated as a matter of regulatory compliance. Abiding by the local safety codes is a condition for operating a business and, as such, necessary, but generic regulations do not cover the specifics of any factory, and are usually not sufficient to make the workplace actually safe. This requires explicit attention as part of work design. Layouts affect the safety of the work environment in at least two ways:

1. **Preventing repetitive stress.** Ergonomically inappropriate tasks may be harmless if done once but cause fatigue or ailments if repeated hundreds of times daily.
2. **Preventing accidents.** Discrete events that disrupt routine operations and make employees unable to work, temporarily or permanently.

Repetitive stress happens for example when an operator works above the shoulder or below the knee, hand-carry heavy parts, or has work positions that require him or her to twist the body. Doing this kind of task once, as in changing a bulb in a ceiling light or squatting to plug a cord into an electrical outlet at floor level at home, is not a problem.

Doing it every minute all day long, on the other hand, makes anyone unnecessarily tired every day and, over a period of months or years, can cause chronic ailments. A century ago, Frank Gilbreth noticed that bricklayers bent down to pick up every brick, so he built a device to present the bricks at work height in an easy to grab fashion.

Layout designers must actively look for accidents waiting to happen. Following are a few usual examples:

- **Fires can start anywhere**, with particularly disastrous consequences where explosions are a risk. It means paying attention to fire codes is the minimum. It may mandate spacing between machines, planned evacuation routes and assembly points, and mandates to park cars facing outwards, but layout designers must look further and, for example, isolate flammable and explosive materials, even if it is not in the code.
- **Traffic accidents.** Sharing space between people and trucks around the plant is an obvious risk but so is sharing space inside the plant between operators and vehicles like forklifts or small trains. That's why their pathways must be as separate as possible and, where physical separation is not possible, movements of vehicles and people restricted to areas respectively marked as transportation aisles or production space.
- **Falling from heights.** The only surface you can't fall from is an even floor. Any elevated platform operators stand on needs a means of keeping them from falling off, either the side of a machine or railings.
- **Trip hazards.** The floor must be even, clear of cables, and free of slick, slippery liquids.
- **Overhead hazards.** Accident risks are not only below the operators' feet; they can also be at head height, in the form of pipes, valves, or other hard objects they can bump into. For operators, this is best avoided by keeping the space around their heads free of any obstacle. Maintenance technicians, who have to crawl into tight spaces, protect their heads with hard hats or bump-caps.
- **Hazards from the work itself.** Operators handle objects that are hot, cold, chemically aggressive, or have sharp edges. The operators are expected to wear personal protection equipment and follow procedures but, quite often, do neither because they feel it slows them down. The challenge of eliminating these hazards without slowing down operators is similar to mistake-proofing.

7.4.6. **Design for well-being**

How people feel impacts their consciousness and alertness. The look and feel of a factory is not to be underestimated. Green areas, daylight, normal temperatures, and fresh air helps people stay awake and have positive effects on their general well being (consider the extreme example in Figure 7.25). More subtle issues such as quick and easy removal of waste and dirt helps keep the work area clean and clear, which reduces the risk for accidents to happen.



Figure 7.25 After a layout project, a small garden with a pavilion is placed where a painting booth stood before in a truck factory in West Virginia, USA

Safety is the basis. Factories should also ensure a level of *well-being* for the people who work there. Besides working in production or support, the employees of a factory enter at the start of their shift and leave at the end, change into and out of work clothes, take breaks, eat and drink, use rest rooms, and attend meetings. Operators enter the plant in street clothes and arrive at their work area in work clothes, where the company provides all the machines, tools, materials, and instructions they need. For this to happen, each operator needs a locker for personal items and, depending on the extent of the change required, a locker room providing sufficient privacy.

At lunch time, if operators walk 5 minutes each way from their work positions to the cafeteria and they stand in line for another 5 minutes, their 30-minute break is, in effect, cut down to 15 minutes. This is a losing proposition: No work is done during 30 minutes but the employees don't have 30 minutes to rest. This is the most fundamental reason to provide "green corners" within feet of each team's production area for members can sit down, eat and drink during breaks, and meet as needed.

While there are general rules about the number of rest rooms that need to be provided for a group of people, it is better to err on the side of having too many than too few, as waiting in line for a stall is not a productive use of employee time. The quality of the rest rooms, also, is revealing of management attitudes. In some cultures, differences in position and status are expressed in every aspect of work life,

including different uniforms, cafeterias, rest rooms, and reserved parking spaces; in more egalitarian ones, hierarchy is limited to the work. In all activities that are not the work employees are paid for, they are all equal, wear the same work clothes, eat together, use the same rest rooms, and have no reserved parking spaces. The best factories provide the same facility quality at the shop floor as in the offices.

Stop and Think 7.5
Is it important that the factory is visually appealing for workers and guests? Why or why not?

Box 7.5 **The Herbert Lindner Factory in Berlin (1933)**

During a study tour of German factories in 1935, Toyota R&D manager Kazuo Kumabe made the following comment about the Lindner factory:

> *"They make many products in buildings set in a huge garden with modern glass windows. In the garden there is a beautiful lawn, a flower bed, carefully tended by a specialist. This is what they think of as a factory. In order to make good precision machines, it is necessary to keep up the spirit of the workers by refreshing and resting their eyes with greenery. They made the garden for this purpose."*

Lindner doesn't make polishing machines anymore but their machines are still used in India. The Herbert Lindner factory was designed by architects Martin Punitzer and Hans Simon in the "new objectivity" style, and built in 1932–1933. The Lindner plant is now a Berlin landmark and featured in the architecture museum at the Technical University. Figure 7.26 shows the plant at the time of Kumabe's visit and today.

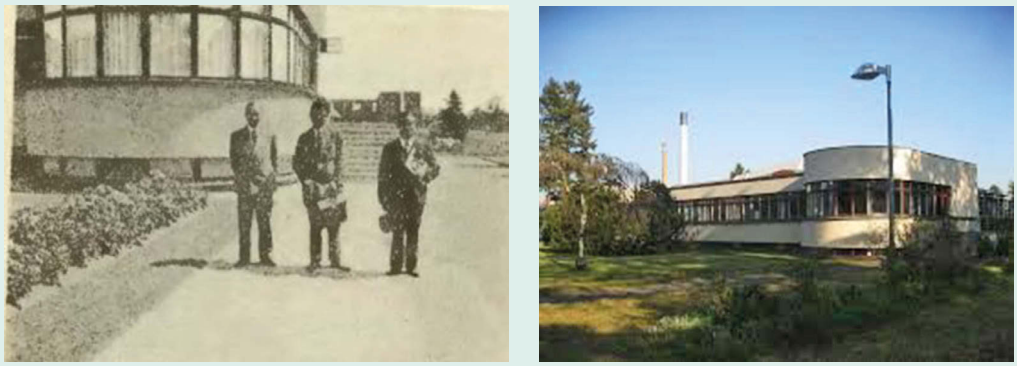


Figure 7.26 K. Kumabe at Lindner factory in 1935 and Lindner factory today

7.5. Monuments

> **Definition**:
> A monument is a machine or structure that is too expensive to move.

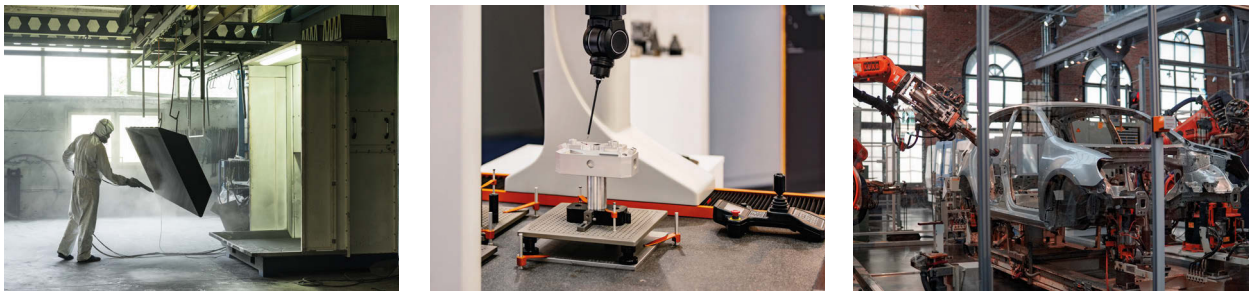
Any machine or structure can be moved if the benefits of doing so are sufficient. A monument, by definition, is a case where management decides it is never justified. Although there are exceptions such as precision optical equipment, monuments are usually larger than the other machines in the shop, and their operation is complex, which means that it is not just loading, processing, and unloading one part.

Monuments frequently have load sizes greater than one, and may take mixed loads of several products at a time. Frequently, they require multiple passes on the same part, and because they are rare and expensive machines, they have to be used on several products. Furthermore, they often require special facilities, such as foundations, exhaust and waste disposal systems, or a higher roof than the rest of the plant.

Examples of machine monuments include the following:

- **Smelting ovens**. Smelting ovens are large, heavy, and wired, which makes them impractical to move.
- **Die casting machines for steel or aluminum**. Not only are such machines large and difficult to move, but they also have a number of characteristics that make them unsuitable for collocation with machining or assembly. They have molten metals brought either in tubs or through pipes, and their output needs to cool down before it can be handled. Operations also involve the manipulation of dies weighing tons and lubricants that are difficult to contain.
- **Rolling mills**. The rolling mills that produce the steel coils that are fed to stamping machines in car plants occupy entire buildings, but even small rolling mills are monuments.
- **Tunnel kilns**. Tunnel kilns move ceramics products slowly on carts through a tunnel organized in zones corresponding to each step of the firing process.
- **Body welding shops in car assembly**. In most cases, plants use monuments when they cannot help it. In the case of car body welding, however, over the past 30 years, automation has turned a manual assembly line process with multiple workstations into a monolithic forest of robots that is in fact a monument. In the early 1980s, the car industry introduced robots to make the key welds that gave the car body its shape, but the numerous intermediate welds needed to give the body strength were still manual. Today, the entire main body welding is automatic, and only small subassemblies are still welded manual.
- **Paint shops**. In low-volume applications, paint booths can sometimes be integrated with other production operations but high volumes, quality, safety, and environmental issues, as well as the special technical skills required, dictate the use of a separate paint shop. In the case of cars, this includes electrocoating of the bodies through immersion, the use of robots to apply the top coat, and handling and ventilation systems to evacuate the paint mist and prevent contamination.
- **Heat treatment**. Many metal parts go through hardening operations in furnace ovens.
- **Autoclaves for large parts**. For curing composite materials, autoclaves are needed. When parts are big, such as the floors, bodies, and wing for jet fighters or commercial aircraft, the autoclave ovens must be big too. Autoclaves are heavy and require a lot of energy, which means that when they are first installed in a factory, they are likely to never move.
- **Plating shops**. Whether electroplating processes require monuments or not depends on the size of the parts. Some aerospace applications clearly do.
- **High-precision machines**. Sometimes machines that are not especially large cannot be easily moved because the precision of the work they do requires the stability provided by a 20 ft.

concrete foundation or adjustments that may require months of engineering work. Examples include optical equipment used in semiconductor processing or in making laser sights, or measurement systems in aerospace manufacture.

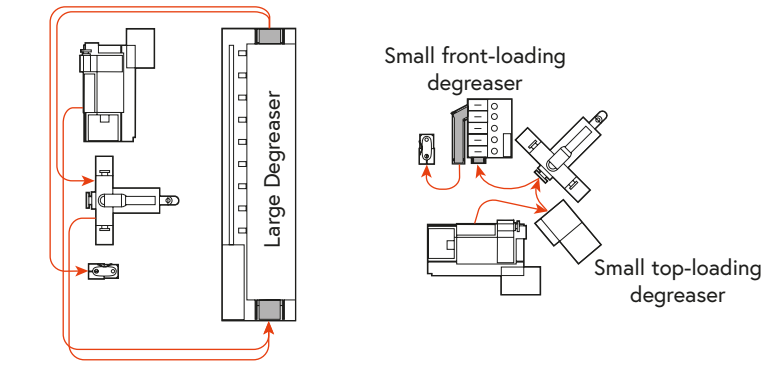


a) Sandblasting booth b) 5-axis measurement system c) Body welding workshop (Credit: BMW)

Figure 7.27 Examples of typical monuments

Technically, a monument can also be a part of the building infrastructure, such as pillars, firewalls, cranes, or special rooms. When these installations hinder flow, they become monuments. In all the preceding examples, the use of monuments is dictated by the *process*, and inevitable without introducing different technology. It is not uncommon, however, to see plants with self-inflicted monuments, such as the following:

- **Flexible Manufacturing Systems (FMS).** FMSs are integrated machines and storage systems. FMSs are often monuments simply due to their size and complexity.
- **Automated Storage and Retrieval Systems (AS/RS).** Storage devices like pallet racks or flow racks can easily be split and relocated as needed, but the same cannot be said of an AS/RS. You cannot split off half of an AS/RS and move it elsewhere in the plant. An AS/RS is sometimes the right technology for a particular application, but the user organization should consider the requirements before committing to making a monument that does nothing but store and retrieve materials.
- **Centralized degreasers.** In machine shops, work pieces are often cleaned and degreased between operations. Many managers think they can save money by centralizing the cleaning function and having a single point at which to deliver cleaning materials and from which to collect waste. As a consequence, however, they break the process flow, make cells impossible, and create a new bottleneck. Degreasing is a secondary operation, and should be integrated with the primary operations that require it. In cells, this means that cleaning or degreasing stations are wedged in between machines (see Figure 7.28).



a. Centralizer degreaser monument b. Local, small degreaser in cell

Figure 7.28 Centralized versus local degreasers

7.5.1. What is wrong with monuments?

There are several reasons to dislike monuments and avoid them if possible:

- **Monuments are expensive to buy.** When there are smaller alternatives, one single monument is frequently more expensive than building several small machines adding up to the same capacity. The labor required for the smaller machines does not have to be higher with multifunction operators in cells, or even with operators who attend to more than one machine concurrently.
- **Monuments are expensive to operate.** For monuments that use cutting tools, such as multi-spindle gantry profilers, tooling costs are also higher because of the need to manage the tools in matching sets. Besides, monuments in general are more difficult to maintain than smaller, simpler machines.
- **Monuments break flow.** From a lean manufacturing point of view, monuments are rogues: They break up the flow of parts in cells by forcing excursions. WIP accumulates in front of them to build up full loads and to wait for each product's turn in the setup sequence. It also accumulates after them, as normal machines work off the full loads released.
- **Monuments are inflexible.** Many monuments, though not all, are inflexible. Multi-spindle profilers do the same work with all spindles, so that, when demand drops, you end up using only a fraction of the spindles. Some other monuments, such as FMSs, do not have this weakness.

> **Stop and Think 7.6**
> Given the many drawbacks, why do firms invest in monuments?

7.5.2. What to do about monuments

Impact on material flows must be on the list of review criteria for new machines. Equipment acquisition decisions made in Facilities departments tend to be based on available space, power and plumbing requirements, waste disposal requirements, and the fire and safety codes in effect. Unless the Production department makes its voice heard and brings this issue up, it may be ignored. As a result, it is likely to get monuments, and the design of in/out buffers for these monuments as well as that of operator jobs will be afterthoughts.

Production engineers must not only bring these issues up, but must carry out the cost of ownership calculations needed to support monument avoidance. Members of the Facilities or Maintenance department, for example, know equipment mostly from the point of view of space and energy requirements or changing filters, not running parts through it. Their perspective, while relevant, is not that of production.

The following subsections provide details on what to do about monuments:

- Avoid them
- Organize logistics to mitigate their upstream and downstream impact
- Locate monuments strategically
- Use prep and finishing cells
- Sequence flow to approximate takt operations
- Reduce process times through monuments
- Reduce setup times of monuments
- Apply visual controls and improve supervisory control

Avoiding monuments

The first strategy with respect to monuments is to avoid them. Monuments often try to accommodate a multiplicity of parts, made at inconsistent paces, and the starting point in an effort to avoid monuments is a review of processing requirements and takt times. The next step is to find equipment of the right size for each part at the pace at which they are to be made. If this equipment is not available off the shelf, then it may be developed in house. While that is not always possible, many companies have had success at it. There are organizational implications to pursuing this strategy, and a specific view of the role of manufacturing engineers as tinkerers.

Alternative technology is sometimes an option, although it has the drawback of requiring numerous approvals, for example if you are making safety parts for automotive or aerospace applications. Examples of alternative technologies that have been used for this purpose include the following:

- Induction hardening to replace heat-treat furnaces
- Sputtering to replace wet plating processes
- Change of material properties to allow gluing instead of welding (consider Box 7.6)

Sometimes the issues are not technical at all, as in the case of an AS/RS. The acquisition of such a monument is best avoided by improving the logistics of parts supply in ways that make it unnecessary.

Box 7.6 **Using composite materials to reduce monuments in car manufacturing**

Founded in 2015, UK-based eco-friendly transportation solutions manufacturer ARRIVAL is using composite materials instead of steel and aluminium for the vehicle body of busses and vans. This has at least three great benefits for manufacturing vis-a-vis traditional manufacturers:

- No need for large stamping factories
- No need for large welding factories, the materials are glued together with industrial-grade glue that is stronger than the materials
- The color is embedded in the composites eliminating the need for a paint booth



Figure 7.29 An ARRIVAL truck (Credit: ARRIVAL/UPS)

Hence, ARRIVAL eliminates the three traditional monuments in car manufacturing: Stamping, body-in-white welding, and painting. This enables ARRIVAL to set up much smaller and cheaper "micro factories" closer to the market, compared to the large automobile factories that still are common.

Instead of an assembly line, ARRIVAL reportedly uses many assembly cells, which allows a higher product mix on the shop floor (different takt times) and maintenance or replacement of a cell without any affect on running operations.

Organize logistics to contain upstream and downstream impact

Sometimes it is infeasible to avoid the acquisition of a monument. Other times you inherit one you cannot immediately get rid of. What do you do? First, you need to recognize that it is a *rogue machine* and contain its influence on the rest of the plant. The following dialogue reveals the impact of failing to do so:

"Why do you run 100-piece lots through final assembly?"

"Because we can't run batches any smaller through the paint shop."

Here, the paint shop is the monument, and the plant's management is allowing it to dictate batch sizes to downstream operations. The norm for the plant should be one-piece flow, and, if a machine won't support it, then we organize logistics around it to quarantine it, as shown in Figure 7.30.

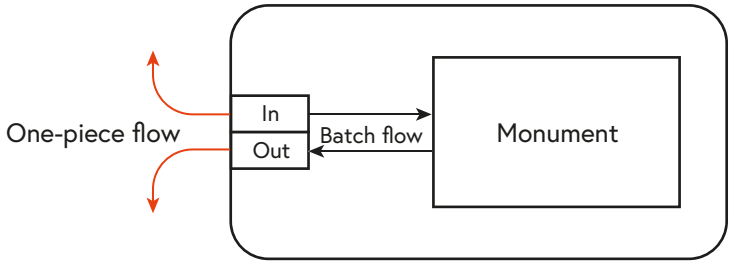


Figure 7.30 Logistics around monuments

Locating monuments strategically

Since monuments, by definition, cannot be moved, their location within the plant is a topic to be addressed at the time the plant is designed. A common pattern is to give monuments a central location within the plant. The rationale for this is that, since all parts pass through the monument, placing it in the center minimizes the average travel distance to and from all other departments. While this logic is valid for locating, for example, a distribution center in a supply chain, it is not applicable to in-plant logistics, where costs are driven by the number of transportation steps rather than the distance traveled. Considering this, there are often more advantages to locating monuments next to an outside wall, for supply of outside materials, waste disposal, and safety.

Using prep and finishing cells

While the processing times through the monument can reach tens of hours, the total times through the monument area usually are far in excess of the processing times as parts wait to be prepared, loaded, unloaded, cleaned, inspected, and transported afterwards. While in the hands of the monument team, parts also sometimes undergo operations that later have to be undone, such as insertion into plastic bags or bubble-pack envelopes from which they are later retrieved. In addition to work on parts, pre- and post-processing often also involves storage, retrieval, handling, assembly, cleaning, and inspection of fixtures, molds, or tools. The elimination of unnecessary steps in pre- and post-processing and the reduction of the waiting times is not one single problem with a point solution, but rather a large number of small inefficiencies that need to be addressed one by one. Cell layouts are often applicable and can help move from batch-and-queue to one-piece-flow in pre- and post-processing.

Sequencing flow to approximate takt operations

The next step is to work on the operation mode of the monument itself, with the objective of having all parts visiting it coming out each on its own takt time. For this purpose, we need to work out the implications on operators and tooling. One useful tool is the manual simulation of the monument's operation as a board game, in which the pieces are operators, product units, pallets, etc. The simulation is used in several ways:

1. Set a repetitive loading sequence leading to making each part within a period of time as close as possible to its takt time.
2. Plan a response for when parts of the machine are down for any reason. The results of the simulation are collected on a whiteboard in the form of either a Gantt chart or an Ybry chart, which is similar to work combination charts.

Reducing process times through monuments

Monuments are often bottlenecks and reducing their process time can be the most direct and cost-effective way to increase the plant's capacity. Such an undertaking requires an engineering task force with the following knowledge and skills:

- The physics, chemistry, or biology of the process in a quantitative way
- The statistical concepts of process capability and statistical Design of Experiments (DoE)
- Machine programming

Some engineers may be able to get on knowing the science of the process alone, but their experiments will take longer than if they had been planned using statistical designs. Conversely, an engineer who knows statistical DoE but has no notion of what the process does will also be limping along and waste time rediscovering what is already known. Finally, the implementation of process changes requires programming the equipment. Most plants do not have the full panoply of skills, and struggle accordingly.

How large the task force is and how long it needs obviously depends on the machine and its improvement targets. Six months to a year is not uncommon for such a project, and the task force needs access to the equipment to run experiments. As a consequence, this kind of project is most difficult to conduct when every available minute of machine time is needed for production. It is best conducted before it is needed, during a business downturn. At such times, not only are the internal resources available, but the suppliers' are as well.

Reducing the setup times of monuments

Because many products go through a monument, the ability to switch quickly between products is key to its flexibility. The methodology for setup time reduction is known as *single-minutes exchange of die* (SMED). "Single-minutes" means "in less than 10 minutes," and "exchange of die" comes from the original application, to automotive stamping presses.

In the 1950s, stamping presses that took two hours to change generated large amounts of WIP when Toyota introduced leveled sequencing in final assembly. While Toyota's competitors in the US and Europe worked on the assumption that setup times were fixed and tried to find workarounds, Toyota's people tried to reduce them to match their mode of operation. The result was SMED, a method that has

been proven many times, not just in stamping presses, but in all sorts of other machines and lines, including both monuments and cells.

On light machines, SMED can sometimes be achieved in a few days but, on monuments, it may take months and require modifications to the machines, for example to allow horizontal die exchange, and die standardization. The reason it is not as widely used as one might expect for such a successful technique is that most managers lack the requisite perseverance. The most common misunderstanding is that the goal is to increase machine efficiency. It is not. When Toyota was reducing setup times on stamping presses, the goal was *not* to increase the total number of stampings made for every hour worked but to increase their variety. The goal of SMED is to increase flexibility, in the sense of supporting a broader mix of products within the same amount of net available production time.

Applying visual controls and improving supervisory control

Monuments often have a control room like the one shown in Figure 7.31, and it is such a common sight that you may not find anything objectionable until you look closer.



Figure 7.31 Heat treatment control room

The following are a few questions worth raising about control rooms:

- Why is a separate room needed? The monument is to the right, and computer controls are on the floor, collocated with machines elsewhere in the plant. In the control room are circuits handling both power distributed for the machines and data that is displayed in a multiplicity of information-poor displays. The data could be routed separately to an information-rich display with, for example, natural mapping, located next to the machines, where operators could attend to fixtures and parts between the times control actions are required from them.
- Why isn't the operator facing the monument? If the operator were facing the monument, he would see both the movements of parts and the status of the equipment through andon lights and direct observation. The control room in Figure 7.31 at least has a glass wall through which the operator can see the equipment. Others have solid walls and windows facing away from the equipment.
- Are we sure we are monitoring the real situation? The coupling between the control room and the equipment also needs to be checked, as there are many cases where the controller's capabilities are not fully used and where a visit to the control room gives a misleading picture of the operation.

7.6. Layout decision-support methods

Because layout decisions are infrequent, manufacturers usually lack the needed skills and experience in doing it well. For bigger investments in new factories, factory extensions or layout redesign, it is therefore wise to hire external expert assistance. It is advisable to choose consultants that specializes on layout projects or factory engineering in general.

To ask useful questions to the external advice, and eventually make the best choice, manufacturers should be familiar with the following methods:

1. Product-quantity analysis
2. Spaghetti diagrams
3. Flow diagrams
4. Production flow analysis
5. Closeness rating analysis
6. Layout mock-ups and sketch-ups
7. Simulation models

Other methods exist too. Box 7.7 presents the VATI analysis used in the Theory of Constraints (TOC).

Box 7.7 VATI analysis: Four generic plant types according to the Goldratt

In his *Theory of Constraints,* Eli Goldratt categorized simple plants as V, A, T, and I:

- A **V-plant** disassembles. You start with one item, like crude oil or an animal carcass, and you end up with multiple products. In a V-plant, the mix of finished goods is determined by the composition of the incoming materials.
- An **A-plan**t assembles. It starts with a base of components converging towards a few products at the top.
- In a **T-plant**, you fabricate a generic product as in an I plant, and then customize or configure it into many end-products. The key challenge is to keep options open by placing the branching point as close as possible to the end of the process.
- An **I-plant** fabricates one incoming item into a product. It comes in as a casting or a raw wafer, and goes out as a transmission case or a wafer with integrated circuits on it. In an I-plant, the availability and performance of equipment at every operation is key.

Draw the material flow of the plant from the bottom of a paper to the top and you will see the plant type. Designing and operating these different types means dealing with different issues. A large, real plant, can be a combination of shops A, V, I, and T shops. The most common combinations are V base with T top, A base with T top, V base with A top, V base with A middle and T top, and both A and V base with T top.

According to the proponents of this simple method, it can help the identify the best practices one can use to leverage or reduce constraints (bottlenecks) and synchronize the flow of materials. The method is helpful for consultants and students trying to get an overview of a factory, but in reality, it will often not return much more than factory workers already know.

7.6.1. Product-quantity analysis

Product-quantity (P-Q) analysis is a simple yet powerful tool for structuring the *volume* of demand. The idea is to group products in volume categories, to serve as a basis for laying out the production floor. Process similarity among products is not ignored, but it is only a secondary criterion, evaluated only among products that are in the same volume category.

In companies that do not practice P-Q analysis, managers may struggle to take advantage of many other known improvement tools because they see their production as "non-repetitive." What they are actually doing is treating all their products as if they were worst cases. P-Q analysis ferrets out the repetitiveness underlying the apparent chaos in the demand and allows the manufacturer to take advantage of it by structuring production lines to match the actual structure of the demand.

P-Q analysis breaks down the product mix in the following three categories:

- **A-products, or "runners."** These are the few products with volumes high enough to warrant dedicated production lines. Most commonly, these 5 to 10 products account for more than 70% of the total volume.
- **B-products, or "repeaters."** These products deserve a dedicated production line by product family, but not individually. The product families are then usually based on technical similarity. There may be 200 B-products, accounting for 25% of the total volume.
- **C-products, or "strangers."** All others are made sporadically, in a generic facility, to order, and without maintaining work-in-process (WIP) for their components. There may be 2000 C-products, accounting for 5% of the volume.

Typically, there may be 5 runners, 200 repeaters, and 2000 strangers. A *Pareto diagram* of volume by product yields an ABC categorization of products as in Figure 7.32. Some texts recommend adding an XYZ categorization, in which the Xs have a stable demand, the Ys have demand fluctuations around a stable mean, and the Zs have a sporadic, uncertain demand.

The XYZ analysis, however, requires more detailed data than the ABC analysis, as well as more work on demand models to set thresholds between categories. In addition, the XYZ categories are not as orthogonal to the ABCs as an ABC/XYZ matrix suggests: Xs are usually runners and Zs are strangers.

P-Q analysis is based on three to six months' worth of production data, which usually is readily extracted from the company's computer systems into an electronic spreadsheet compatible format like "comma-separated values" (CSV). On occasion, it is a much more difficult challenge. In the worst case, it can be a list of invoices.

In practice, a complete P-Q analysis requires drawing multiple charts as shown in Figure 7.33. The analysis in terms of sales revenues can be misleading. The most extreme distortion appears in industries where the highest numbers of units produced are for promotional giveaways, which generate no income but may still be among your A-products. For this and other reasons, a P-Q analysis is to be based *on numbers of units*, not revenue. You will find that the dominant items are not always the same when you consider different categories. Three usual categories are:

- Sales revenue
- Number of units produced
- Number of components produced

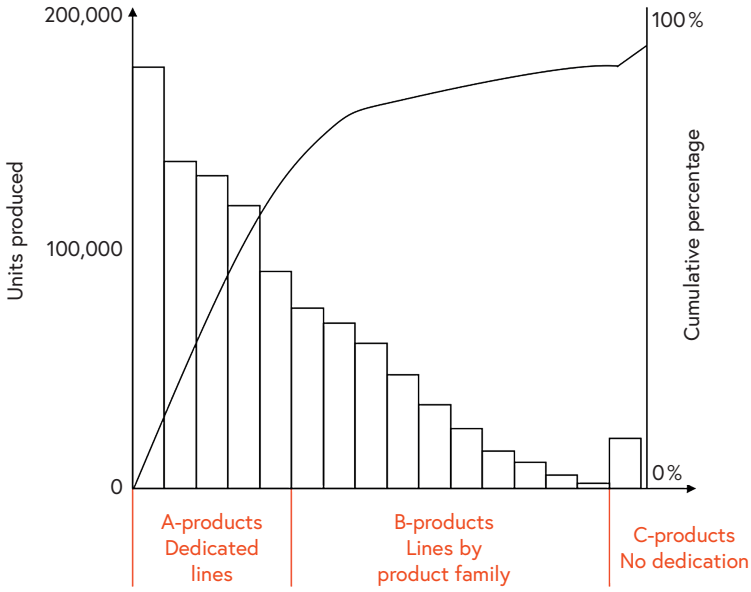


Figure 7.32 Example of the result of a P-Q analysis

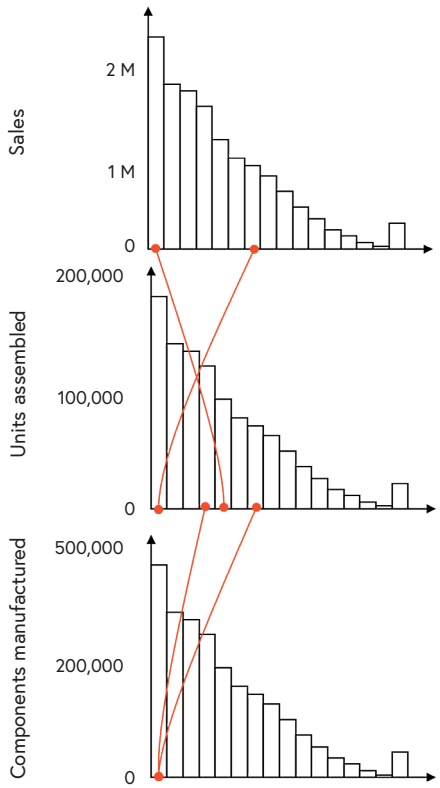


Figure 7.33 Multi-layered P-Q analysis

7.6.2. **Spaghetti diagrams**

Spaghetti diagrams are layout drawings with flows of materials drawn on them. These charts are called "spaghetti diagrams" because of their tangled appearance when finished. Drawing spaghetti diagrams is a straightforward exercise. For every product, follow the path between operations from raw materials to finished product. The finished diagram provides a visual perspective on congestion, busy and non-busy resources, and a feeling of the flow efficiency in the factory.

Figure 7.34 shows an example of a spaghetti diagram used in a real factory planning project. Here the engineers have simply printed a micro-layout of the area they want to improve and used pins and colored rubber bands to draw movement patterns for a typical product, for example the product with the highest volumes. An alternative is to draw lines with a pencil or use drawing software.

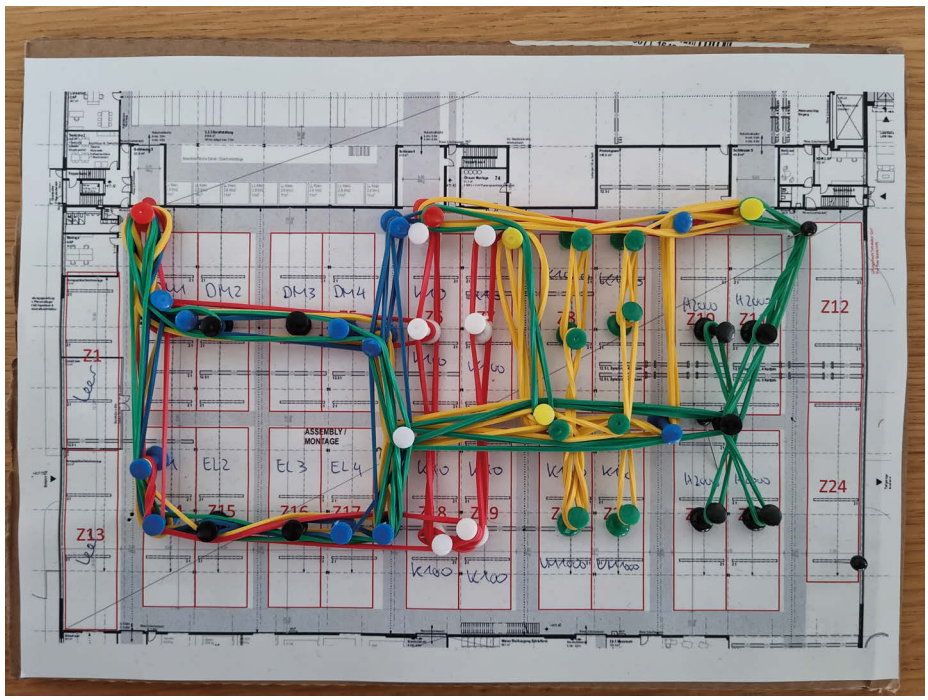


Figure 7.34 A spaghetti diagram for a factory making CNC machines (Credit: Andrea Calfisch)

7.6.3. **Flow diagrams**

Flow diagrams are needed to show operations proceeding in parallel, converging and diverging at particular points. It is an abstract representation, with a limited graphic vocabulary based on, for example, the American Society of Mechanical Engineers (ASME) symbols, with circles for processes, inverted triangles for stores, square boxes for inspection or test, and arrows for material flows. Besides the ASME symbols, there is the Japanese standard JIS-Z-8206, a German DIN standard, an ISO standard, and others. Toyota has its own, which contains additional icons for logistical and information flow patterns specific to lean manufacturing

By showing nothing but the flows of materials, and collapsing large organizations like final assembly or body-in-white assembly each into a single bubble, it highlights what the plant as a whole does. While

the level of detail in process flow diagrams is high enough to address production control issues, it is too low to support the design of assembly stations or assembly lines. This level of detail, which is included in assembly master tables, is missing in these process maps Figure 7.35 shows high-level material flows in a car assembly plant of Mazda, which summarizes in one page the work of about 5,000 people.

7.6.4. Closeness rating analysis

Richard Muther[1] introduced the *closeness rating analysis* (CRA) as a macro-level tool to analyze the need to collocate departments. Departments that should be close may for example share equipment, have operations that follow each other in sequence, or frequently exchange data. Departments that should be kept apart from each other are, for example, a lab with delicate equipment and a department that has equipment with strong vibrations, the Sales department and the noisiest Machining department, or clean zones and dirty machining processes.

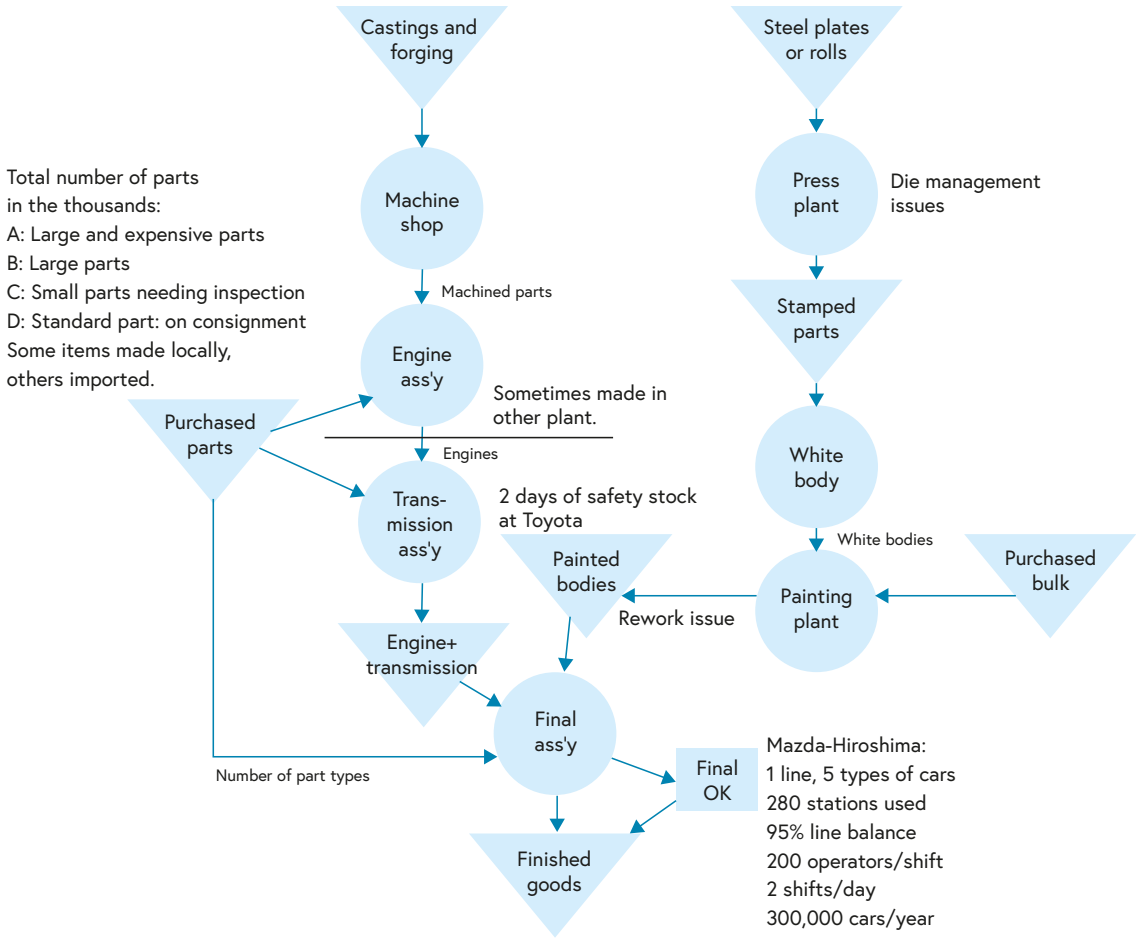


Figure 7.35 Flows of materials in a car assembly plant

1 Muther, R. (1973). *Systematic layout planning*. Cahners Books.

The CRA is a qualitative tool that uses 6 letters for describing the need for closeness between departments, as shown in the example of a jam manufacturers in Figure 7.36 below:

- A= Absolutely necessary
- E= Essentially important
- I= Important
- O= Ordinary closeness
- U= Unimportant
- X= Undesirable

When conducting a CRA, you encounter employees who are frustrated from having to interact multiple times daily with colleagues from another department located at the other end of the building and on a different floor. You also meet others who are equally frustrated while trying to concentrate on technical problems next to customer service agents working on the phone and celebrating each success by ringing a bell. Desirable and undesirable collocation are easy to tell apart.

The distinction between "absolutely necessary" and "essentially important," on the other hand, is more subtle. To make it, you need to clarify two issues:

1. The consequences of the choice on the layout design.
2. The list of criteria to assign a value.

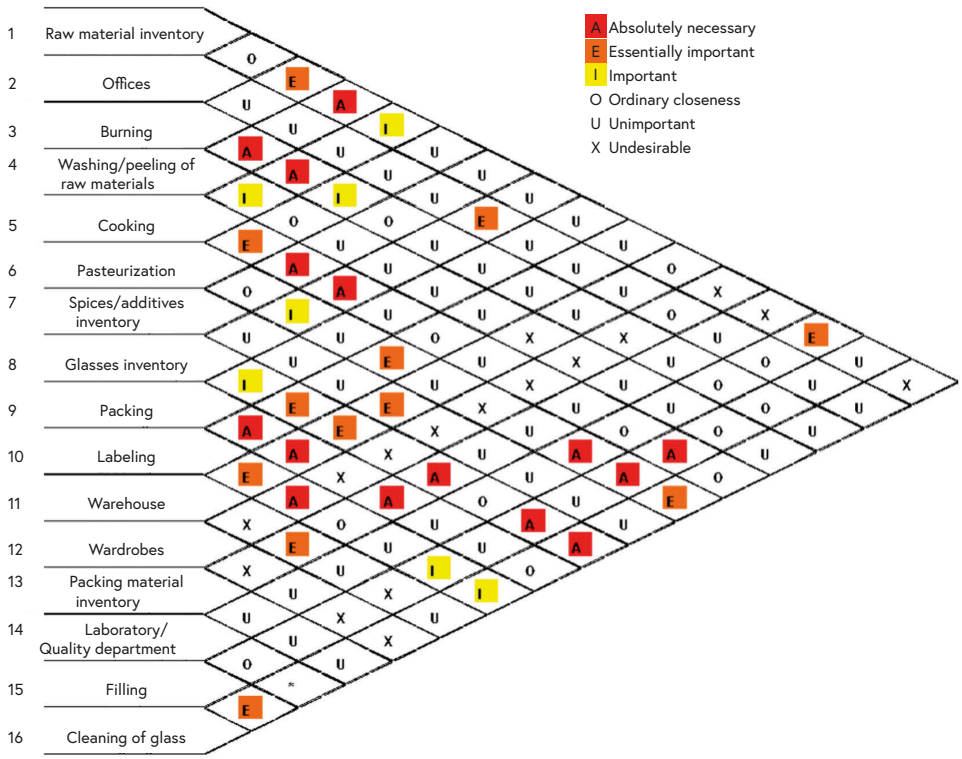


Figure 7.36 Example of CRA for a jam manufacturing factory

Stop and Think 7.7

Check if you agree with the CRA analysis done in Figure 7.36. Make assumptions where necessary.

The extent of interactions with other departments that a department needs to function is called its *coupling*; what its various tasks have in common is called its *cohesion*. A CRA with many colored cells is a sign of high coupling and low cohesion, Rather than just collocating departments in a new layout, the more appropriate response may be to modify their roles.

The CRA is a manual technique from the 1970s. Where companies record the flow of work between departments as database transactions, you can apply *Process Mining*[2] to automatically identify relationships and quantify the volume of work they entail.

7.6.5. Group Technology for "repeaters"

Group Technology (GT) is the grouping of products or components by feature similarity for the purpose of producing them in a mixed-flow line, and is relevant for *repeaters*. *Runners*, by definition, have dedicated lines and *strangers* have a demand that is too low and sporadic. In auto parts, for example, a door panel for this year's model may be made at a takt time of 30 seconds. A similar panel for an obsolete model may be needed at a rate of 300 units/year, with orders at random intervals, and it makes no sense to burden the line making the current model with it.

The breakdown of products into families may begin at a higher level. A manufacturer of components for multiple assemblers of consumer goods may have customers with different requirements:

- Customer *C1* allows them to lay out their lines as they see fit, as long as each operation is performed to their specs.
- Customer *C2* insists on approving every layout change
- Customer *C3* keeps a supplier support engineer in residence.

And all three customers may be concerned about intellectual property leaking to competitors. The manufacturer then operates a focused factory for each customer, within each of which, there are runners, repeaters, and strangers. Within each focused factory, the intermediate level of *repeaters* holds the products or components you can group into families to which you can dedicate lines. GT is a method that is then usable to organize these families.

GT predates database systems and relied on encoding features into numeric codes from the Opitz system as seen in Figure 7.37. Database technology has rendered these cryptic codes obsolete, which part of the GT literature still had not realized by 2016.[3] You can now analyze similarities directly from lists of feature names and values attached to items.

2 van der Aalst, W. (2016). *Process mining: Data science in action*. Springer.

3 Zehtaban, L. & Roller, D. (2016). Automated rule-based system for Opitz feature recognition and code generation from STEP. *Computer-Aided Design & Applications, 13*(3), 1-11.

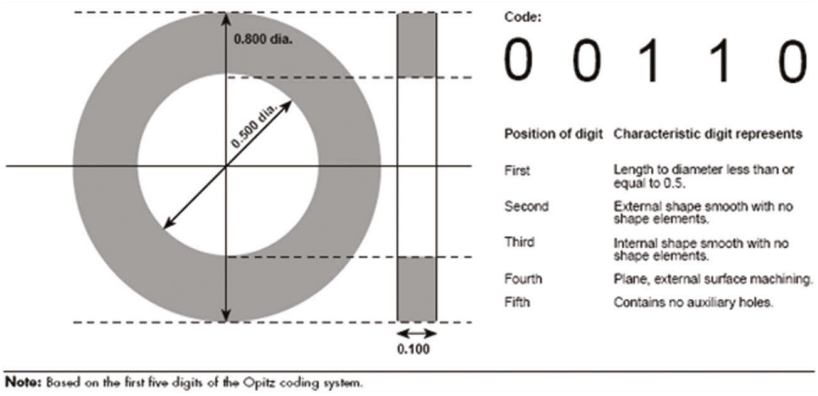


Figure 7.37 The obsolete approach of encoding features

In the 1970s and 80s, John Burbidge[4] developed a technique for turning job-shops into flow-shops using group technology. His method of *production flow analysis* (PFA) is used worldwide and independent of industries and factory sizes. PFA clusters parts that follow the similar processes into part groups or "families," which are produced in dedicated lines.

PFA follows a two-step approach. First, part groups must be identified. A part family is a collection of parts that have similar features – geometric shape, size, material – and are made with similar processes. Identifying part families is a time consuming job done by trained personnel. In PFA, part families are identified by sorting the parts and processes into a two-by-two matrix as seen in Figure 7.38.

PFA chart

Process \ Parts	A	B	C	D	E	F	G	H	I
1		x				x			x
2	x				x				
3	x				x		x		
4				x				x	
5		x							x
6							x		
7			x	x				x	

PFA cluster analysis

Process \ Parts	G	E	A	I	F	B	D	C	H
3	x	x	x	Part group 1					
2		x	x						
6	x		x						
1				x	x	x	Part group 2		
5				x		x			
7							x	x	x
4				Part group 3			x		x

Figure 7.38 PFA part group identification

4 Burbidge, J. L. (1989). *Production flow analysis*. Oxford University Press.

In the second step, production machines must be rearranged into cells. Figure 7.39 shows an illustration of a layout change for the example in Figure 7.38. The layout changes from a spaghetti flow through multiple departments in the left figure to the cell layout in the right figure. Although this example is helpful to explain the concept, the reality is more complex and users of PFA have to simplify to get ahead.

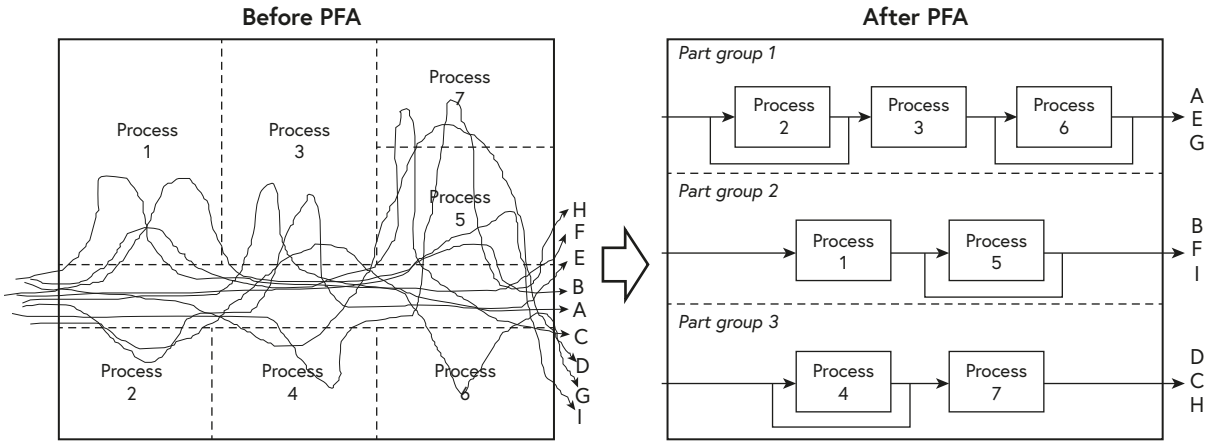


Figure 7.39 Before and after PFA: From job-shop to flow-shop layout

Layout mock-ups and sketch-ups

A variety of three-dimensional models of lines have been used, some on a reduced scale as architects do, others full size, with mock-ups of stations and machines. Reduced-size models are primarily used to explain a finished facility to visitors. As a design tool, generating and maintaining them requires too much work for the additional value they bring compared to drawings.

Full-size models of machines, tables and tools can be built out of corrugated cardboard and tape. They allow operators to move through the "to-be" work environment, develop a sense of what the work experience will be, and provide feedback. Full-size models are the ultimate manual simulation tool. The construction of such a model shows a commitment to usability that is appreciated by operators. Some manufacturing companies have used this "cardboard engineering" technique for years, using mock-ups detailed enough to actually assemble parts and thereby test and debug the process. The cardboard stations may be replaced several times as they wear out, during an exercise that may last two days. Then the feedback is incorporated in the design and the line is built for real. This approach is useful for reasonably small products in short assembly processes. It is difficult to imagine how it could be used, for example, in final assembly of cars or heavy industry factories.

Figure 7.40 shows an example of cardboard mock-ups being used to plan a production line layout for a new generation mechanical heart-pumps in the Thoratec Corporation factory in Zurich, Switzerland. The people in the photo are operators and managers of the current production line. The participants are evaluating different layouts by shifting the cardboard tables around, discussing and taking notes on the flipcharts to the left. The boxes marked with blue papers are roughly the size of tabletop ovens used in curing operations.



Figure 7.40 A cardboard mock-up of a layout project in Thoratec Corporation, Zurich, 2018 (Credit: Lukas Gratwohl/Abbot)

Small-scale models can be built with LEGO or other materials. It gives workers an immediate feeling of the future layout. As these models will never be 100% accurate, they are most helpful during the lay-out-planning phase. Machine avatars can easily be moved and colored thread can be used to show flow of different products.

In line layout design, as in architecture, digital 3D images have been displacing physical models. One tool used for this purpose is SketchUp, which illustrates lines and full factories before installation. Digital 3D models allow you to "walk through" the building and lines, and zoom in on specific details. It gives a good understanding of what the environment will look like. The biggest limitation with digital mock-ups is that they cannot simulate moving production. For that, companies must use simulation models.

Virtual Reality (VR) systems are under development for line and station design. They allow the designer to preview the working experience of operators with different body sizes. The virtual parts they "handle" have locations to reach for but are weightless. With a given geometry, VR cannot yet make a difference between holding a part weighing 1 oz or 5 lbs.

7.6.6. **Simulation models**

Discrete-event simulators are software programs that simulate the state changes of a system as a result of events and planned responses. Originally applied to computer networks, they have also been used over the past 30 years on manufacturing operations, where the events are, for example, the start of an operation in a machine or the completion of that operation. Their main output is *performance measurements* on simulated operations that serve to validate a design. Some can also produce *animations* that are used to show management a proposed line in operation.

Today's most advanced tools are overkill in simple manufacturing processes, where a full-size mock-up can be built in a fraction of the time and cost it takes to develop a computer simulation. These systems,

however, have found a niche in complicated assembly systems involving many products and the coordination of multiple, independently controlled automatic systems, such as the flexible assembly of car bodies with welding robots and automatic guided vehicles (AGVs) for part supply.

Examples of companies that have groups specialized in studying multi-million dollar systems using simulation include Airbus, AkzoNobel, Boeing, Honda, Ford, General Motors, Harley-Davidson, Renault, Siemens, Toyota, and IBM, among others. The attitudes of individual plant managers vary, but some do withhold approval of projects until validated by simulations. Smaller companies may not be able to afford an internal simulation group. Many suppliers of materials handling systems, however, offer simulation services to customers.

As an example, consider Siemens Tecnomatix Plant Simulation software (Figure 7.41), which is among the more advanced commercially available simulation software packages. It enables the simulation, visualization, analysis, and optimization of production systems and their related logistics processes. According to Siemens, the software can assist the optimization of material flow, resource utilization and logistics for all levels of plant planning, from global facilities and local plants to specific production lines. The software provides 2D and 3D visualization and animation.

A simulation model does more than help plan a new factory. Perhaps its biggest benefit is the opportunity to run experiments and what-if scenarios without disturbing an existing production system. This however, would require keeping the digital model updated when any changes are introduced to the shop floor.

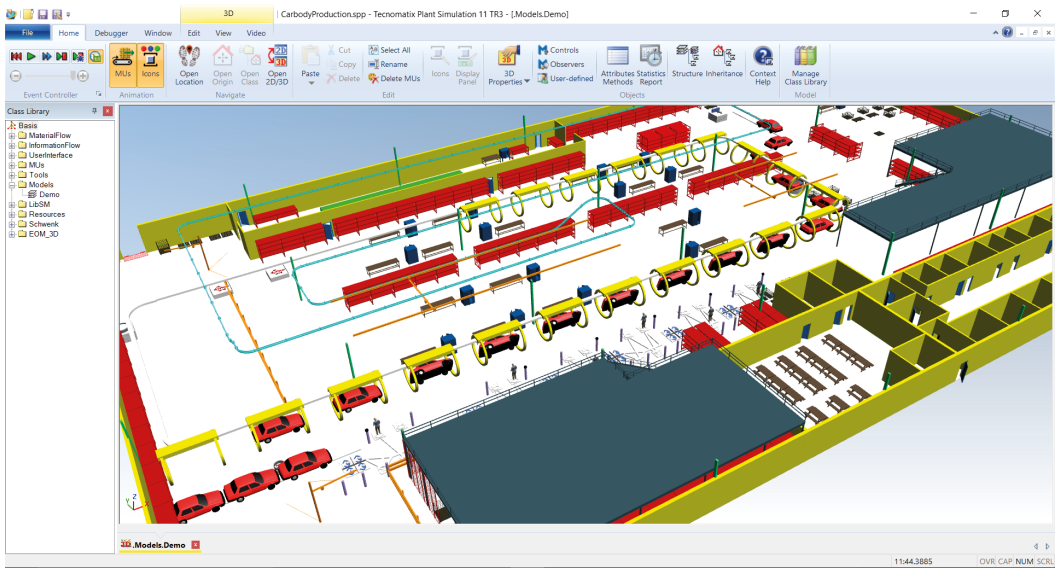


Figure 7.41 User interface example of Siemens Tecnomatix Plant Simulation (Credit: Siemens AG)

Summary

The most essentials points covered in this chapter were:

1. The factory layout is the physical arrangement of departments, workstations, equipment, storage areas, and aisles within a manufacturing firm's facility.
2. Layout design is a strategic decision.
3. The type of layout must match the characteristics of the products in terms of size, weight, and shape, demand volume and variation, safety and risk, the technology and skillset employed, and the need for customization and flexibility.
4. There are three generic layout types:
    - Fixed-position layout
    - Job-shops
    - Flow-shops
        - Cells
        - Assembly lines
        - Continuous flow
5. A factory should be planned inside-out.
6. There are many special requirements in layout planning that requires attention.
7. There are four key phases in a layout project:
    - Pre-planning
    - Macro layout: Ideal and realistic
    - Micro layout
    - Installation
8. A monument is a machine that cannot be moved at a reasonable cost. Monuments are often expensive rogue machines that break one-piece flow, require complex tooling, and are difficult to maintain.
9. Safety and well-being must be considered in the layout.
10. Some decision-support methods are particularly helpful for the factory planner:
    - P-Q analysis
    - Spaghetti diagrams
    - Flow diagrams
    - Production flow analysis
    - Closeness rating analysis
    - Layout mock-ups and sketch-ups
    - Simulation models

Test yourself

1. Why are layout decisions important?
2. What is the difference between a job-shop and a flow-shop?
3. What type of generic layout would you suggest for the following products?
    a. Mechanical door locks
    b. Bottled water
    c. Space rocket
    d. Chewing gum
    e. Custom furniture
4. What are common mistakes in layout projects?
5. What are "green corners" and why do they exist?
6. What is a monument?
7. If your factory has one autoclave that all products must pass through, and it only takes batches of 10 pieces per cycle, should you let 10 pieces be the standard batch in your factory? Why or why not?
8. What is cardboard engineering and when can it be used to plan layouts?

Further reading

Benjaafar, S., Heragu, S. S., & Irani, S. A. (2002). Next generation factory layouts: Research challenges and recent progress. *Interfaces*, 32(6), 58–76.

Hyer, N. & Wemmerlöv, U. (2001). *Reorganizing the factory: Competing through cellular manufacturing*. CRC Press.

McGinnis, L. F., White, J. A., & Francis, R. L. (1992). *Facility layout and location: An analytical approach*. Prentice Hall.

Muther, R. (1973). *Systematic Layout Planning*. Cahners Books.

Nelson, W., Amundsen, A. E., Lee, Q., & Tuttle, H. (1997). *Facilities and workplace design: An illustrated guide*. Engineering & Management Press.

Phillips, E. J. (1997). *Manufacturing plant layout: Fundamentals and fine points of optimum facility design*. Society of Manufacturing Engineers.

Schenk, M., Wirth, S., & Müller, E. (2010). *Factory planning manual: Situation-driven production facility planning*.

Stephens, M. P. (2019). *Manufacturing facilities design & material handling* (6th ed.). Purdue University Press.

Sule, D. R. (2008). *Manufacturing facilities: Location, planning, and design* (3rd ed.). Taylor & Francis.

Wiendahl, H., Reichardt, J., & Nyhuis, P. (2015). *Handbook factory planning and design*. Springer.

Part 3

Making Information Flow

DOI:10.4324/9781351110310-10



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

8 IT and OT in Manufacturing

Programmability is the most sweeping change to have occurred in the late 20th century in the way humans interact with machines. And not only in manufacturing, but in most other industries and in everyday life. In manufacturing, Operational Technology (OT) covers the systems that drive, monitor, and respond to machines, robots, conveyor systems, or vehicles; Information Technology (IT) covers those that only exchange data with other systems or with people.

In this book, "IT/OT" refers to the continuum of software and hardware used, from sensors on the shop floor to analytics for CEOs. We use it because it is descriptive of all the electronic technology used to handle information, and it is non-judgemental. It carries no implication that this technology is either a blessing or a curse. What we are discussing in this chapter is the impact of programmability on manufacturing operations, from the level of what happens inside production machines, including operator-machine interactions and the programming of the machines, to high-level support systems such as ERP and cloud services.

Some manufacturers see IT/OT as key to competitiveness in modern manufacturing and a driver of productivity improvement but many have yet to realize its potential. IT/OT has transformed the work of managers, engineers, maintenance technicians, and production operators. In production engineering, it has spawned the new job specialty of machine programmer, and created in the programs themselves a new class of assets for manufacturing companies to manage.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Separate and explain different types of IT and OT systems in manufacturing.
2. Establish the IT infrastructure needed in a factory.
3. Understand how data are shared across IT/OT systems.
4. Understand the role of databases in manufacturing.
5. Understand how data are shared across IT/OT systems.
6. Understand the basic building block of industrial OT; the PLC.
7. Explain the role and function of key manufacturing IT/OT systems: CNC, IoT, SCADA, MES, and ERP.
8. Understand how IT is moving towards apps delivered as software-as-a-service from platform-based clouds.
9. Overcome challenges related to IT/OT implementation in manufacturing.
10. Understand the pervasiveness and limitations of desktop applications such as slideware and spreadsheets in manufacturing.

DOI:10.4324/9781351110310-11

8.1. The manufacturing IT/OT stack

The most common summary of the IT/OT used in an organization is as a *technology stack*. It summarizes complex interactions by piling each layer of functions on top of a lower layer that it invokes and exchanges data with. Below each layer is its *infrastructure*; above it, its *superstructure*. If, as in Figure 8.1, you pry open a stack, you expose a tangled web of connections, and hiding it is the point of the stack diagram.

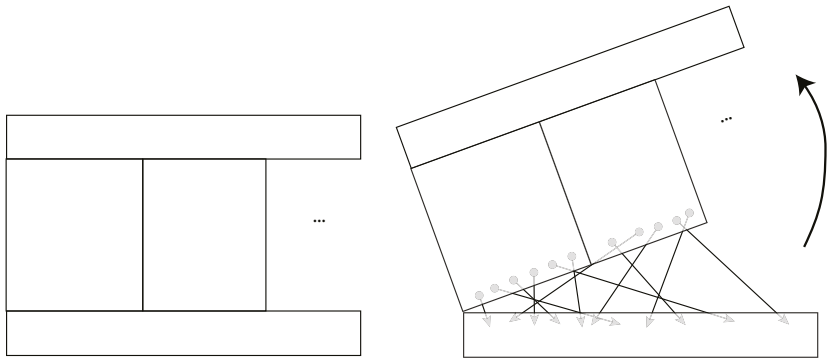


Figure 8.1 Hiding complex interactions by stacking

Figure 8.2 shows the typical IT/OT stack of a manufacturing company. The IT systems at the foundation concerns communications among the systems above. The OT systems in the middle layers interact directly with production machinery and with operators. The IT systems at the top interact with engineers and managers. The stack uses many acronyms that may be unfamiliar to you:

- **5G** 5th generation mobile network: A new global wireless standard after 1G to 4G networks. With superior data transfer capabilities, it enables mass-adoption of IoT and cyber-physical production systems.
- **AI** Artificial intelligence: AI is the simulation of human intelligence processes by machines. As used here as part of supervisory process control, AI may be applied to diagnosis of failures or to local scheduling. This is not the only application of AI in manufacturing.
- **AM** Additive manufacturing: Construction of a three-dimenisonal part directly from a digital model.
- **APS** Advanced Planning and Scheduling: Generation of actual sequences of production activities over time, possibly with algorithms from Operations Research (OR) or artificial intelligence (AI).
- **Auto-ID** Automatic identification: Barcodes, data matrix, and sensors technologies that captures materials, vehilces, and machines in real-time.
- **CAD** Computer-aided Design: Software used mostly by product designers but also for layout descisions by factory engineers.
- **CMMS** Computer-aided Maintenance Management System: Scheduling, transaction processing, and analysis of maintenance.
- **CNC** Computer Numerical Control: Special-purpose computer that directly controls a machine.
- **CRM** Customer Relationship Management: Manage communications with current and potential customers.
- **ERP** Enterprise Resource Planning: Originally used for high-level planning, today mostly commercial transaction processing.

- **HMI** Human–Machine Interface: In general, an HMI is any interface between a computer and a human. In this context, it is more specifically about interfaces between operators and production machinery.
- **HR** Human Resources: This system maintains, with security, all the data about employees and contractors, job descriptions, and work histories.
- **IIoT** Industrial Internet of Things: Parts, machines, and other resources that are connected online to a computer network via sensors, instruments, and applications.
- **MES** Manufacturing Execution System: Shop floor transaction processing and related functions.
- **PLC** Programmable Logic Controller: Special-purpose computer that switches motors on and off, opens and closes valves, and communicates the status of a production line to humans.
- **PLM** Product Lifecycle Management: Revision management for product specs.
- **QA** Quality Assurance: Software that helps comply with external mandates and solve quality problems.
- **SCADA** Supervisory Control and Data Analysis: Intermediaries between operators and one or more machine controllers.
- **WMS** Warehouse Management System: Directed putaway, picking, retrieval of materials in a warehouse.

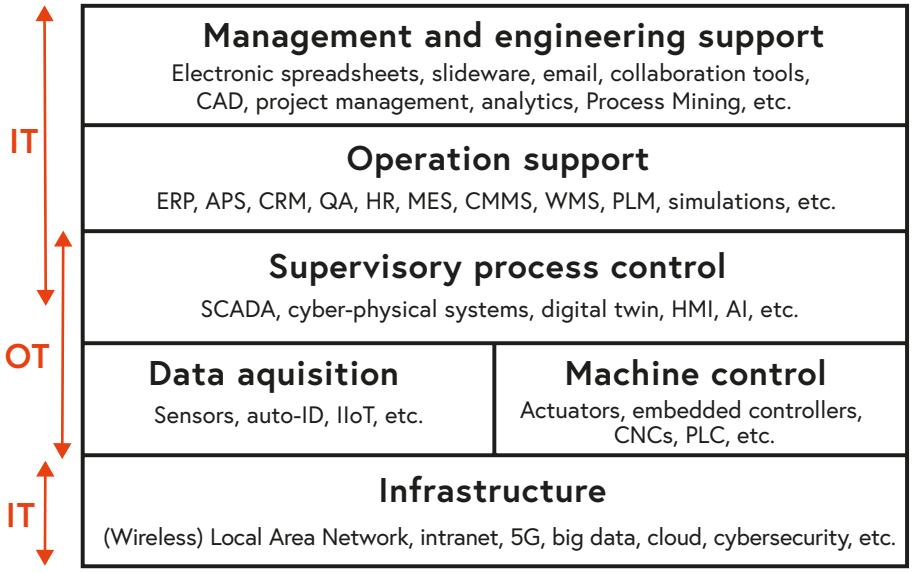


Figure 8.2 A manufacturing IT/OT stack

The OT system are close to machines and processes and use technology that is almost only understood by process engineers or industrial engineers. The IT systems in the upper layers, on the other hand, are technically similar to what you would find in a bank or an insurance company. They use generic hardware, operating systems, and communication networks. These systems are usually operated by an IT department populated with engineers who know these generic tools but not manufacturing. The most commonly used IT tools in manufacturing remain to this day Microsoft Excel and PowerPoint.

The red arrows to the left refer to the departments in charge of the systems. Many manufacturing companies have an IT department that operates networks and servers, and provisions employees with computers and apps but doesn't have expertise in control of production machinery. For this reason, OT is selected, implemented, and maintained by manufacturing or process engineers. The boundary between the layers usually operated by Engineering/OT and IT is fuzzy and their collaboration limited.

At the extremes, the distinction between IT and OT is clear: The embedded controller of an ion implanter is OT and the corporate production planner is IT. In between, the boundary between the two is fuzzy, and drawn differently in different companies. Where OT ends and IT starts is often the object of conflicts between departments, that have consequences on how well OT and IT play together. This is why the IT/OT arrows in Figure 8.2 has has overlap between IT and OT, often fought over and hotly contested.

The "Industry 4.0" approach seeks to eliminate the distinction between OT and IT. Since OT is at the *edge* where data and the physical world interact and IT has been migrating to the *cloud*, this has been called *edge-to-cloud* integration and *cyber-physical production systems*. The challenge of making it happen, however, is not technical but human. It encompasses a broad range of skills, and communities of users with different priorities and willingness to share data.

Box 8.1 **The ISA95 standard of enterprise integration**

In this chapter we use the IT/OT stack shown in Figure 8.2. There are alternative models. Without endorsement, we include one of the most common standards in case you encounter it elsewhere: An international standard provided by the International Society of Automation illustrates data integration in the enterprise in the form of a pyramid. The ANSI/ISA-95 model Enterprise-Control System Integration model. It separates five levels of industrial IT as shown in Figure 8.3. This model seeks to be an information model that defines IT systems by architecture.

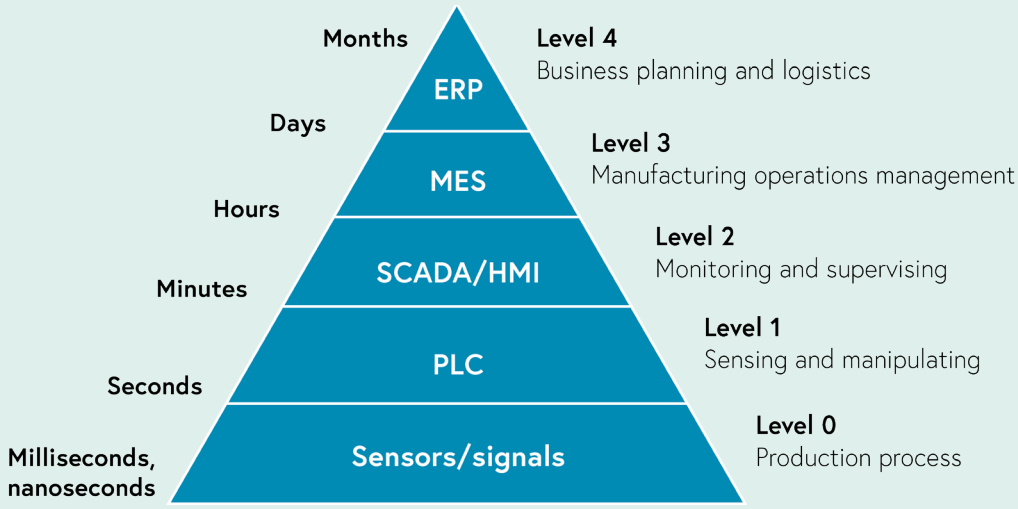


Figure 8.3 The ISA-95 standard separates five levels of IT

8.2. Infrastructure

End-users do not see all the infrastructure of IT/OT, yet it enables everything they see. It is also subject to rapid change and a domain that calls for vigilant technology scouting – yet, knowing that the latest technology is not necessarily appropriate for a given factory. In local networking, for example, the state of the art in the 2020s is no longer Wi-Fi but 5G. It should be a consideration when building a new factory but replacing Wi-Fi with 5G in an existing factory may not be justified.

The architecture of the systems used in manufacturing is a reflection both of the IT/OT available and of specific needs. If manufacturers could buy all-in-one systems providing all the functions in Figure 8.2, the sharing of data between these functions would be the system supplier's problem. As different departments find specialized systems to better meet their needs than any all-in-one system, the manufacturer needs to worry about data sharing between every pair of layers in the stack.

8.2.1. From client/server to publish-and-subscribe

In business IT, dumb terminals sending keystrokes to mainframes faded away in the 1980s as more and more employees were provided with networked personal computers at their desks. Applications could then offload user-facing tasks like collecting input at displaying output to the personal computer, send packaged requests to software running on another computer shared with other users, and receive responses. This became the *client/server model* and it is pervasive today.

Client and *server* are software concepts but "server" has come to be used to refer to the hardware they run on. Computers are sold as "servers," and companies operating on the web are said to have "server farms." In fact, it's not hardware that makes them servers but software, with names like *Apache* or *Windows Server*. You can have clients and servers running on the same computer where, for example, a word processor acts as a client to a print server. Multi-user database systems like *MySQL* or *ORACLE* run as servers with multiple clients that may all be on the same machine.

In manufacturing, different departments often have their own systems, each with an *Application Programming Interface* (API) that allows the others to treat it like a server and this is where trouble happens. The "P" in API stands for "Programming" and, for the accounting system to retrieve work-in-process status from the MES, it needs a program invoking the API. The same is true for the MES to check equipment conditions from the CMMS, or the list of operators qualified to fill in for an absentee from HR, and so on. Hence, if you have 5 systems, you need 10 interfaces between pairs. Every time a software supplier upgrades any of the systems, you need to upgrade 4 interfaces.

This problem was solved in the late 1980s by a different mode of interaction: The *publish-and-subscribe model* invented by Dale Skeen. Instead of interacting with each other, each system exchanges data with what was then called a "middleware" layer but could as accurately be called a local cloud. Each system *publishes* messages to this cloud and *subscribes* to messages about particular topics. It's like sending tweets with hashtags and declaring your interest in following particular hashtags, except that it is done by software rather than people and it's formal, so that each system can use the data it subscribes to in executing transactions or queries. *Extensible Markup Language* (XML), *JavaScript Object Notation* (JSON), or *Yet Another Markup Language* (YAML) are usable formats for this kind of message.

In publish-and-subscribe, upgrading one system only requires an upgrade to its interface with the middleware layer, not to all the other systems. This was the key motivation in the adoption of this model by Intel in the early 1990s. Cloud providers offer publish-and-subscribe services for *application-to-application* (A2A) and *application-to-person* (A2P) communication.

In the OT domain, *Message Queuing Telemetry Transport* (MQTT) is a protocol used for this purpose in the Internet of Things (IoT), in which the service that matches subscriptions and publications is called a *Broker*. Figure 8.4 is an example of a sequence of message exchanges in MQTT. Mosquitto is open-source software implementing MQTT.

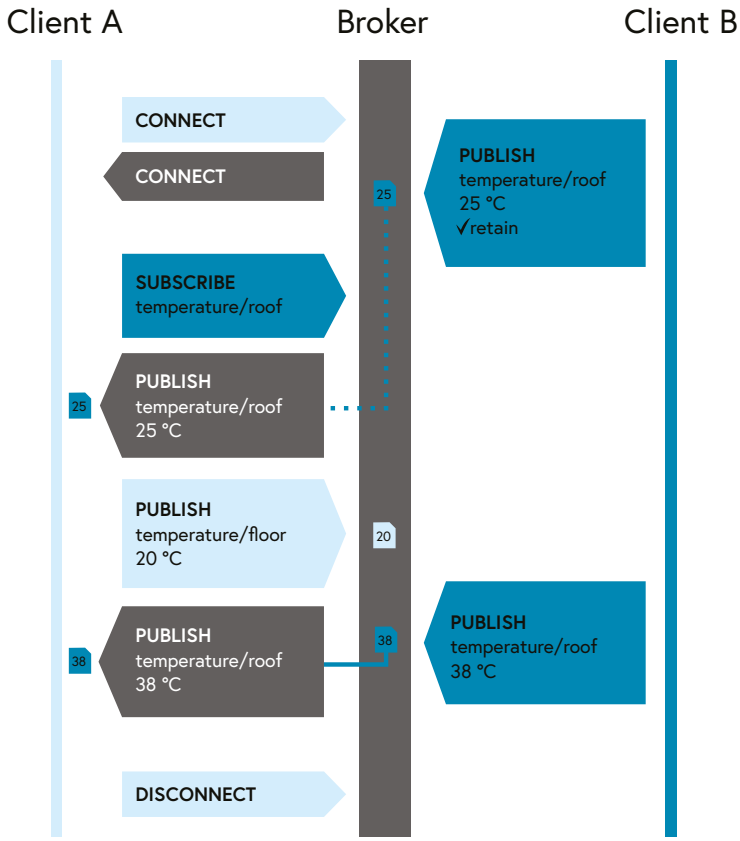


Figure 8.4 Publish-and-subscribe example with MQTT (Credit: Simon A. Eugster, Wikimedia CC BY-SA 4.0)

A system in a publish-and-subscribe network does not have to share *all* its data. It only needs to publish what is necessary. Box 8.2 shows three examples of application of this model.

> **Stop and Think 8.1**
> Why don't manufacturers use one integrated IT/OT system instead of a multitude of different systems from different vendors?

Box 8.2 **Examples of publish-and-subscribe networks in use**

ACCESSORIES AT DEALERSHIPS. Construction companies buy excavators from dealers, with buckets that are specific to the work they want to do and are sold as accessories. What happens when a dealer does not have the desired bucket in stock? The fastest way to get one is from another dealer but a dealer does not have access to other dealers' stocks. The dealer's system can publish the demand for this bucket. Other dealers' systems that subscribe to messages about bucket demand pick up this message and respond with a quote on price and delivery conditions. The original dealer can then choose the best offer. The problem is solved without the dealers having to share their inventory data.

DYNAMIC SCHEDULING OF PARALLEL LINES. Assume a factory with multiple lines having overlapping capabilities, of different vintages and with different modes of operation. Each line has a SCADA system that is used to schedule it. The plant's production schedule includes an order that requires a part that can be made in any of these lines. How do you decide which line to use? The planning system can *publish* a request for bids from the different lines. The SCADA systems of each line *subscribe* to messages on such topics and each one quotes back a delivery time. The planning system then assigns production of the part to the bid with the earliest time.

TEMPERATURE READINGS. A machine controller reads a thermometer every 1/100th of a second but publishes only one every second. The high-frequency readings are useful to the controller for early warning of accidents. During routine operations, however, the temperature does not change abruptly and there is no technical analysis conducted at a higher level than the machine that requires more than one reading per minute. One per second is therefore amply sufficient.

8.2.2. The need for repositories of shared data

There is more to data sharing than exchanging messages, as some of the shared data needs to endure as part of a shared environment. If data sharing is limited to messaging, then the plant as a whole has no memory other than the data maintained by each participating system. Likewise, a team that exchanges project data by email has all its project data in the email systems of the participants and no shared repository. Collaboration software such as Slack, Yammer, and Sharepoint instead keeps all the project data in one place for all participants to retrieve and update – assuming it is used with discipline.

An engineer who publishes a query about the failures of the equipment used to make a particular product will receive separate responses from the maintenance systeme (CMMS) with all the equipment failures, and from manufacturing execution (MES) with the product history. This engineer will then have to cross-reference and filter the results to get the query precisely answered, and will only be able to do so to the extent that a partial models of the plant across the systems are consistent. If, for example, the CMMS and the MES use different IDs for equipment, joining data from the two systems may be practically impossible.

This implies that the manufacturing organization centrally maintains the *master data* for all items used by more than one system. The details needed to read a pressure from a sensor may be private to a controller but the name of the process chamber where the readings are taken, their unit of measure, accuracy, timing, etc. are shared with every other system involved with posting vital signs of the process or responding to anomalies.

In any operating system, there are the following four categories of data:

- **Master data** – timeless. This includes product and process definitions, bills or materials, routings, equipment, facilities, job descriptions, etc. It is definition data that is timeless in the sense that it remains valid and effective until management decides to change it.
- **Planning data** – the future. This is data about the future, including yearly capacity plans, monthly sales and operations plans, daily production plans, down to the schedule of what each resource on the shop floor should do next. It changes with the passage of time.
- **Status data** – the present. This is data about the current state of production and resources. Which machines are processing what, and are they on schedule, ahead or behind? Which ones have alarms or are down? Who is at work and where? This is real-time data. The "present" value of any parameter is usually taken to be as of its last update.
- **History** – the past. As events occur, like shipments to customers, receipts from suppliers, operation completions, quality problem reports, machine failures, etc. data is kept in history for traceability and analysis.

8.2.3. Types of databases

The tools that computer scientists have developed to help users share data are *databases, not spreadsheets*. Databases are under the hood of e-commerce sites and social media. At work, the ERP, MES, SCADA, and other systems also use databases in a way that is also transparent to end users. Even for a single user, there are database systems that provide ways of organizing, storing and retrieving data that are richer and less error-prone than spreadsheets. These database systems are included in the professional version of the best selling office suites but most professionals don't know how to use them.

From small to large, databases fall in the following categories:

1. Personal, lookup databases
2. Individual databases used for analytics
3. Shared databases used for transaction processing and analytics
4. Distributed, shared databases

An individual database system is meant for a single user who needs to work with data, as opposed to the larger databases that are for many users to read and update the same data in a coordinated fashion. Single-user database systems are local apps on their users' computers with each database as a local file. The most common ones are Access on Windows systems and FileMaker on Macs. They are an improvement over spreadsheets for local data storage and retrieval but not about data sharing and security.

The multi-user databases used by SCADA, MES, or ERP systems are servers over the network that can be simultaneously queried and updated by multiple clients. The clients may be the single-user database systems or other apps.

The separation between the server and clients does not add complexity for end users but it does for the IT/OT staff. The server lets multiple users work on the same data at the same time, using elaborate protocols to prevent inconsistent updates. These capabilities are commonly referred to as "ACID" (Atomicity, Consistency, Isolation, and Durability). Manufacturing professionals don't need to know how the servers do it, but they need to know that they do it, understand the value of it, and insist on using only ACID-compliant databases.

In the 2020s, Oracle, SQL-Server, and MySQL are commonly used multi-user relational database management systems (DBMS), relational meaning that they work exclusively with tables. Mongo DB and Couchbase are more recent DBMSs, known as "NoSQL," meaning that they work with nested lists of name-value pairs, as described in Section 8.2.2. The latest concept is the graph database, in which relationships are objects in their own right, as opposed to being defined only in terms of matching values in table columns as in the relational model.

Data warehouses

To solve problems, engineers and managers need access to historical data from multiple systems. *Data warehouses* are special kinds of databases designed to facilitate this retrieval. Data warehouses concentrate historical data from multiple systems. They are updated periodically, most commonly daily, by Extract-Transfer-Load (ETL) processes that retrieve the recent history from, for example, ERP, MES, SCADA, or CMMS systems, and restructure it into a common information model. Managers and engineers can then pull clean, complete, and consistent datasets for analysis, as shown in Figure 8.5.



Figure 8.5 Data warehouse

In a factory that has a well-run data warehouse, users can instantly retrieve datasets that would otherwise take days to assemble. Database vendors usually provide data warehouse solutions as applications of their database management systems (DBMS) products.

Data lakes

The ETL process of a data warehouse can substantially transform the data from multiple sources to fit within a common information model, but it is overkill for data that is unlikely to be needed. Contracts, for example, contain many clauses that never need to be retrieved unless a conflict arises. A project team may analyze a short segment of the video recording of an operation as part of an improvement project but the entire recording is kept for possible future use.

This kind of data needs to be kept in case it is needed but it doesn't need to be processed until it is. The idea of the data lake is to index the data but keep its content in its original form and throw it as is into the "data lake," and wrangle it into a usable form only as needed. Data lakes are often stored in the cloud through services like Microsoft's Azure Data Lake or Amazon's S3. The logic is illustrated in Figure 8.6.



Figure 8.6 Data lake

In-memory databases

In-memory databases are used for status data that needs to mirror the actual situation of the plant. The amount of data is small but rapidly accessible by multiple apps. For the current value of any parameter of the factory's activity, users don't wait until an update is published. Instead, they want the last value published, kept in the in-memory database. This is the plant-wide equivalent of the digital twin of an operation maintained by a cyber-physical system.

In-memory databases enable minimal response times by eliminating the need to access disks. They are ideal for real-time analytics or applications that require minimal response time. In manufacturing, an example is an andon call on an assembly line or a real-time visual quality control process. In-memory databases used in manufacturing include SAP's HANA and TIBCO's ActiveSpaces.

Cloud computing companies separate between "hot" and "cold" data. Hot data must be available immediately when needed and can be stored in an in-memory database. Cold data are less time-critical and can be stored on disks as in a data warehouse or data lake.

Document databases

Master data, subject to revision management, require yet another type of repository called a *document database*. Like data lakes, document databases hold data in a variety of formats but are focused on the creation, retrieval, update, and deletion (CRUD) of documents like product specs, work instructions, training videos, and contracts with suppliers. Unlike data lakes, these documents are subject to revision management, and the right version must be retrieved in situations like product recalls.

Big data

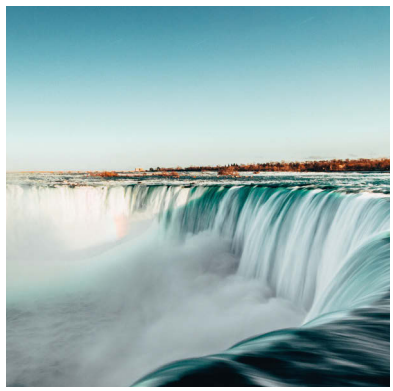
What constitutes "big data" changes over time and varies by industry. How "big" is manufacturing data? Figure 8.7 illustrates the volumes of manufacturing data before the computer age and today with the volumes generated by activities like e-commerce.



Old factory data



Modern factory data



E-commerce

Figure 8.7 Comparative data volumes across old and new factories versus modern e-commerce

Until the late 20th century, manual data collection on ledgers and paper spreadsheets was the norm. The volume was a trickle of data. In today's factories automatic data acquisition systems and IIoT devices produce the output of a firehose but this is still tiny when compared with the flood of data coming out of companies like Google, Amazon, or Facebook.

Nomenclature, process and product definitions, equipment, suppliers, people, inventory status, and history in a factory are unlikely to exceed a few hundred gigabytes. User clicks on a large website, by contrast, can generate multiple terabytes/day. Even packing the shop floor with sensors capturing multiple parameters at higher frequencies than needed and monitoring the entire floor with video cameras will not come close to this kind of data volume. The volume of data produced by a factory must be considered in choosing the hardware and software configuration, as well as the kinds of analytics that can be used.

The statistical methods of the 20th century were developed for testing and estimation with small samples, on the order of 10 data points, and other, simpler ones for large samples, but "large" may mean as few as 100 data points. This was, for example, the kind of data sets that the old Statistical Process Control (SPC) of the 1920s was designed for: Small sets, analyzed manually. Today, these methods are subpar to data analytics. The techniques available to slice, dice, visualize, summarize, and draw well

supported conclusions from large datasets have exploded along with the computing power needed to use them.

As an infrastructure for distributed storage of big data, many web companies rely on Apache's *Hadoop* and *MapReduce*, respectively, for distributed data storage and for analytical processing. These tools are *not needed* in companies that don't have this level of data volume.

8.2.4. **Manufacturing cybersecurity**

While IT/OT security and data security are real issues in manufacturing, publications on manufacturing cybersecurity are at the early research level at this point. That does not stop cyber-criminals from disrupting manufacturing companies and asking ransom money for their webpages or IT systems. With the increasing use of IT in manufacturing, particularly cloud-based solutions, cybersecurity risks increase drastically. Manufacturing is a significant target for cyber-criminals.

Manufacturers are facing multiple threats:

1. Identity theft
2. Phishing
3. Spear phishing
4. Spam
5. Compromised web pages
6. Compromised IT infrastructure

The lack of data security is among the technical barriers to data sharing. The challenge is to provide solutions that are effective yet add no friction in the flow of work. Banks make the login process of online banking ever more complex. Besides a username and password, it now involves a one-time security code received by phone, correctly typing a Captcha, counting the number of bicycles in a grid of photographs, or answering personal security questions. Customers accept it because they understand the need for proper authentication and they only go through this process occasionally. In manufacturing, you want the same level of security *without* adding labor.

One thing is to invest in security systems and backup solitons to reduce the likelihood and effect of a cyber-attack. Another issue is what to do once you have fallen victim to it. Depending on what systems and data are compromised, many companies choose to pay part or whole of the ransom in order to be able to restore critical operations or avoid public loss of face. Most cyber attacks never make it to the news because the companies fear repercussions from stake holders. Box 8.3 reports a radically different response made by aluminum company Norsk Hydro when it was attacked in 2019.

In the quest for increasing digitization of manufacturing, a question worth asking for managers is, "In the case of a cyber-attack, could we run our factory *without* our latest software?" Many companies – like Norsk Hydro – can, but only because they have experienced operators, engineers, and managers who already did it before. The ability to run a factory in an analog or off-line mode as a Plan B is worth maintaining. It requires a strategy, as does cybersecurity.

Box 8.3 **Cyber attack on Norsk Hydro**

On March 19, 2019, the aluminum giant Norsk Hydro fell victim to an extensive cyber attack. The attack affected Hydro's global organization, including the forced closure of entire factories in the extrusion division and the considerable reduction of operations in others. In many factories, IT systems had to be shut down, and operations could only be kept running by returning to old-days analog procedures and workarounds.



Figure 8.8 Hydro factory in Portland, US, shifted to manual operations in March, 2019 (Credit: Norsk Hydro)

The cyber attack was made possible when one of the 35,000 employees unknowingly opened an infected trojan email from a trusted customer (spear phishing). Hackers invaded the IT systems and planted ransomware named LockerGoga. The ransomware encrypted the files on desktops, laptops, and servers and posted a ransom note on the screens of the corrupted computers: "Greetings! There was a significant flaw in the security system of your company... We exclusively have decryption software for your situation...."

At the first emergency meeting, Hydro executives decided that they would not pay the ransom and be fully open about the attack. Hydro slowly resumed operations by rebuilding the IT infrastructure and relying on data from backup servers.

Hydro has documented the total cost of the cyber attack to be around USD 70 million, which was covered by insurance. None of it went to the cyber-criminals. In 2021, Europol detained 12 individuals in Ukraine and Switzerland who were charged with disrupting the businesses of large corporations with cyber-attacks, including the Hydro attack.

8.2.5. Data sharing in manufacturing

When it comes to IT/OT, the first concern should be to have systems that actually help the users; the second, that these systems play together well. It implies, in particular, that they share some of their data.

Data is retrieved from multiple systems for many purposes:

- **Lookup.** For example, you have the Product-ID of a work piece and the ID of the next operation in the MES. These keys serve to retrieve the process program from another database and download it to the machine controller.
- **Search.** For example, used for failure analysis, finding the cause of an equipment breakdown, or locating a misplaced item in the warehouse.
- **Analysis.** For example, used for predictive maintenance, problem solving, establishing process capability, or assessing trends and cycles in product demand.

If search is about finding a needle in a haystack, analysis is about estimating the average number of needles in a haystack, as illustrated in Figure 8.9.



Figure 8.9 Uses of shared data

The challenge is that the main obstacles are not technical but human. The technology to enable sharing has been around for decades (see Box 8.4). Software suppliers have been promoting integration for decades. They sold solutions after solutions to generation after generation of top managers but did not usually convince the end-users.

So how do you get end-users to *want* to share data? If they don't have to share data to get their jobs done, they are generally not interested in sharing. But, if colleagues need data from them to do their jobs, they are usually willing to provide them. What is most needed to promote data sharing is to systematically analyze what data each employee *needs from others* and for what purposes. This analysis commonly reveals cases of employees producing data *nobody* uses and not actually knowing what others need. You can then generalize from individuals to job functions and build a model of information flows within a company or its supply chain. Technology can then reduce the friction in these flows.

> **Stop and Think 8.2**
> What are some possible reasons why human workers are reluctant to share data from processes they are controlling?

Box 8.4 **Data sharing 1980s style**

In the 1980s, MRP systems often ran on IBM midrange computers like the 4381, while MESs ran on DEC's VAX/VMS systems, widely believed to be incapable of sharing data with IBM machines. In factories, the "interface" between the two was an operator reading a screen from one system and manually entering data into the other.

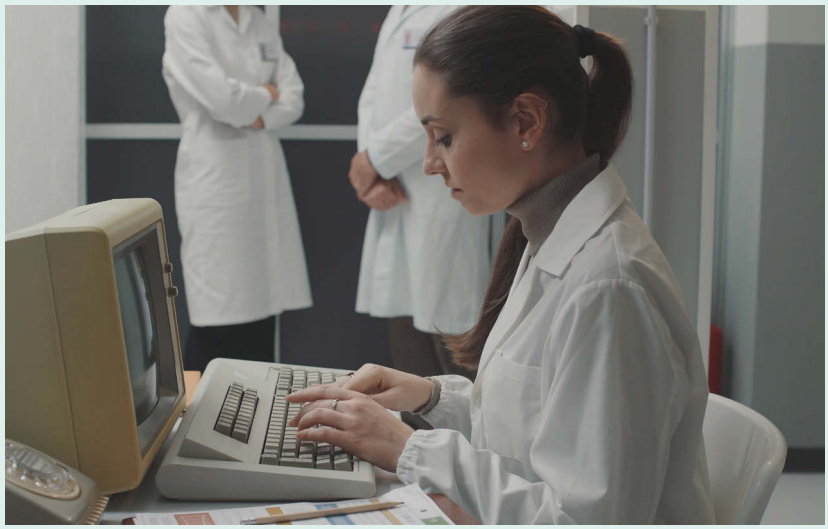


Figure 8.10 Manual transfer of data between IT systems still happens in manufacturing today

In reality, all it took for the systems to share data was an interface box with both plugged in. The real reason the two systems didn't communicate is that the MES users did not want to share their data with the MRP system.

8.2.6. Role of standards

The most advanced protocols are only useful to the extent that the entities sharing data agree to use them. They need to be published as standards that all participants can easily access, learn, and apply. The economic and political aspects of standardization, however, are difficult. In addition, in electronic communications, standards are subject to disruption by innovation.

If standards were the way to promote computer networking, we would be communicating over a worldwide ISO/OSI network, not the internet. Shortly after this massive, 7-layer stack of protocols was published in 1984, the internet took over, for better or worse, and the ISO/OSI model was forgotten, except for some vocabulary. In current documents, you still find terms like “transport layer,” that are legacies of ISO/OSI.

In manufacturing, successful standards often originate within a company, as a solution to a problem. Management publishes it when realizing that it is economically more advantageous to share it with competitors than to keep it secret. It worked, for example, with the *semiconductor equipment communication standard* (SECS), but perhaps because HP, while a respected technology company, was not a

dominant player in semiconductors. When a dominant company proposes a standard, the others often perceive it as a power play and organize to make it fail. Having standards managed by industry groups or professional societies makes them easier to adopt.

To this date, government and international standards organizations have not had a major role in manufacturing data sharing. An open question is whether this has to change for large-scale so-called "digital transformation" of manufacturing. The vision of "Industry 4.0" or "Smart Manufacturing" is to connect all processes in a cyber-physical production system, which should enable descriptive, predictive, and prescriptive analysis of manufacturing processes. For this to happen, the various technologies and processes that need to communicate with each other have to talk the same language or use translating technologies. In the "Industry 4.0" literature, this language is the "OPC United Architecture" (OPC-UA) discussed in Box 8.5.

Box 8.5 **The OPC-UA standard for "Industry 4.0"**

OPC-UA is an open-source, cross-platform machine-to-machine communication architecture developed by the OPC Foundation and was released in 2006. The OPC-UA standard is defined in the International Electrotechnical Commission (IEC) standard IEC62541.

The objective of OPC-UA is to integrate data from heterogeneous systems on the shop floor (PLCs, IoT, mobile devices, etc.) using one common reference language. OPC-UA is the successor of OPC, which could only run on Microsoft Windows and used the obsolete Windows COM communication architecture. OPC-UA is a much more complete architecture that uses sophisticated object modeling, can be certified and is independent of platforms and transport modes. OPC-UA is cross-platform because it works with many programming languages, including Python, Java, C/C++, .NET and others, and can run on different hardware platforms (PC, PLC, cloud-based servers, etc.) and operating systems (Microsoft Windows, Andorid, Linux, Apple OSX, etc.).

OPC-UA is the recommended standard in the "Reference Architecture Model for Industrie 4.0" (RAMI 4.0) developed by the German Electrical and Electronic Manufacturers' Association (ZVEI) to support Industry 4.0 initiatives. It has become the de facto standard for the automation market and "Industry 4.0" technology suppliers. OPC-UA is adopted and offered by many PLC, machine, and device manufacturers. It supports both the client-server model (via TCP and HTTPS) and the publish-subscribe model (via UDP, AMQP, and MQTT).

A limitation of OPC-UA is that it tries to be everything for everyone. It makes the protocol unnecessarily complex and heavy for many use cases. An alternative to OPC-UA is to only use the simpler MQTT. The future will tell whether OPC-UA will become the only worldwide standard for machine-to-machine integration, or if alternatives will arise. Meanwhile, we have to deal with competing standards and opinions.

8.2.7. Communications between systems in supply chains

To share data, machines, systems, and supply chains need a common language. Driven by need, some industries have developed protocols that have been successful. For example, the German Automobile Industry Association (VDA) has been promoting *Electronic Data Interchange* (EDI) for commercial transactions between auto parts and car manufacturers since the 1980s, using its own standard. The VDA standard, however, has not been widely adopted outside of Germany. Many other standards prevail for commerce in general in the rest of Europe, North America, and in the rest of the world. The UN has the EDIFACT standard, and SAP has its own IDOC standard.

EDI messages are now exchanged over the internet, and many of the standards use XML messages. The inspiration for XML is the HTML system of tags that is used to format web pages. Most browsers have a "Show page source" command that lets you see the HTML version of any web page. That an EDI standard uses XML only means that it uses a system of tags, and these systems vary between standards. Box 8.6 shows examples of EDI codes and Figure 8.11 a simple illustration of EDI.

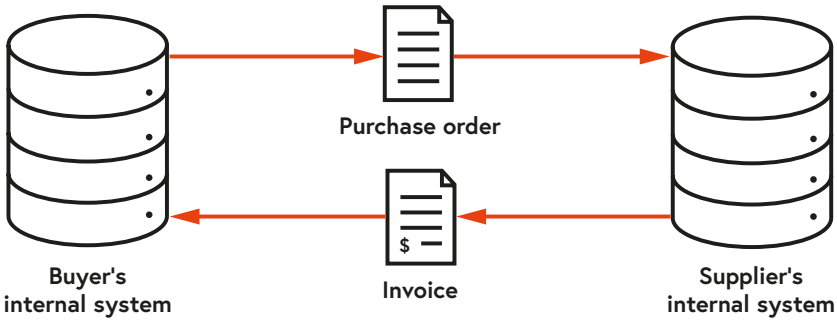


Figure 8.11 The EDI standard enables automatic transfer and revival of data

EDI replaces documents like quotations, orders, invoices, advance shipping notices, etc., with electronic equivalents. Senders generate and recipients process these messages as they see fit, with their own systems. EDI only enables their transmission. The promise of EDI is common message formats but the proliferation of standards has prevented this promise being kept. If you are an auto parts supplier to carmakers in Germany, elsewhere in Europe, in the US, and in Brazil, you may need to master four different ones.

This is not the way consumers buy products online. You access an online store through an app or a web page. Then your clicks invoke functions in the seller's system and pass on your data to these functions. This interaction is specific to the seller. For business-to-business (B2B) transactions, suppliers can provide *Application Programming Interfaces* (API) for customers' systems to interact with theirs as human consumers do with Amazon, eBay, or other online stores.

Assume a customer's system needs to order replenishment of a bulk item because its stock has crossed its reorder point. With EDI, the customer sends a message; with the supplier's API, the customer places the order in the supplier's system by remotely executing the transaction. The API approach seems appropriate when the supplier and the customer are in a long-term, committed relationship but not if the customer would like to elicit bids from multiple suppliers. In this case, the customer can *publish* a request for quotes as an EDI message for suppliers who subscribe to such requests to respond each in their own way.

Box 8.6 **Examples of communication standards: XML, JSON, and YAML**

Following is an example of an XML message concerning a flower for sale in a retail shop:

```
<PLANT>
      <COMMON>Bloodroot</COMMON>
      <BOTANICAL>Sanguinaria canadensis</BOTANICAL>
      <ZONE>4</ZONE>
      <LIGHT>Mostly Shady</LIGHT>
      <PRICE>$2.44</PRICE>
      <AVAILABILITY>031520</AVAILABILITY>
</PLANT>
```

This simple example is human-readable because it's small. Large scale examples, with extensive nesting, are difficult for humans to read. Indentation serves to enhance human readability but is not otherwise necessary. All the values are between tags of the form <name> and </name>.

XML is now an incumbent as an EDI format, and difficult to dislodge. Some companies, however, translate the XML messages into the JavaScript Object Notation (JSON) format that is easier for humans to read and is the storage format in database systems like MongoDB or Couchbase used by major web services. The example above translates as follows in JSON:

```
{"PLANT":
      {"COMMON":"Bloodroot",
      "BOTANICAL":"Sanguinaria canadensis",
      "ZONE":"4",
      "LIGHT":"Mostly Shady",
      "PRICE":"$2.44",
      "AVAILABILITY":"031520"}}
```

In JSON as in XML, indentation plays no role with programs. In JSON messages, structure is expressed by punctuation: braces, quotation marks, semi-colons, or commas; in XML, by tags. YAML is another format that makes messages even easier for people to read by using indentation to express structure. YAML is commonly used to express user settings or preferences in systems but not for EDI. A downside of indentation is that it may be lost when messages are passed around between systems. In YAML, the above example is as follows:

```
PLANT:
      COMMON:           Bloodroot
      BOTANICAL:        Sanguinaria canadensis
      ZONE:             4
      LIGHT:            Mostly Shady
      PRICE:            $2.44
      AVAILABILITY:     031520
```

8.3. **Data acquisition**

Many companies rely on operators to type product item numbers. Numbering systems inherited from the paper-and-pencil age have produced 10 or even 15-digit item numbers with information encoded in them, causing similar products to have IDs that differ only by 1 or 2 digits. In these circumstances, a human being under pressure typing in such IDs is an error waiting to happen, and these errors can have catastrophic consequences. With high-value products like semiconductors, for example, running the wrong program on a diffusion furnace can destroy $250K worth of chips on silicon wafers.

Automatic identification (auto-ID) prevents this type of error. In semiconductors, this was achieved first by using special wafer carriers called standard mechanical interface (SMIF) boxes, which, in addition to protecting wafers from contamination, carried their product and status information in electronic form, and docked into stockers through which that information could pass to the machine controllers. The need for docking has since been removed by using RFID.

In other industries, the most common form of auto-ID is the *barcode*. The products arrive at the machine bearing barcode labels that the operator reads either by moving a handheld reader over the label or moving the parts under a stationary reader. While barcodes are effective at preventing the use of wrong programs, they have the following limitations:

- They require a clean, well-lighted environment and high contrast. Where these conditions cannot be achieved, they are sometimes replaced with ID-squares, also known as *2D-barcodes* or *data matrix* that can be directly laser engraved on metal parts and read through an oil mist.
- Reading a barcode requires labor, and often more than one attempt. Fully automated barcode reading exists on a few manufacturing lines, but it is an exception.

RFID technology promises auto-ID with no operator intervention, 100% reading success on the first attempt, and looser environmental requirements. So far, this promise has yet to be fulfilled and applications in manufacturing have been limited for the following reasons:

- Barcode systems have been successful and therefore not easy to dislodge, especially with a more expensive alternative.
- RFID implementation requires an understanding of electromagnetics that is not widespread among manufacturing engineers. Everyone can understand that a barcode reader must clearly see the barcode, but the conditions under which low-power RF and microwaves are transmitted successfully are less straightforward.
- RFID standards are still evolving, making potential users wary of being stuck with systems that won't be supported five to ten years from now.

> **Stop and Think 8.3**
> Give examples of data that are typically useful to collect from the manufacturing floor. What can these data be used for? How and how long should a company store them?

8.3.1. The Industrial Internet of Things (IIoT)

Over the years, the reach of the internet has grown from mainframes and minicomputers to workstations, desktops, laptops tablets, and smartphones, all of which are computers. With the *Internet of Things* (IoT) and the *Industrial IoT* (IIoT), it extends further, not just to smart devices, like computer controlled production machines and vehicles but also to tags attached to moving objects that can respond to a contactless reader and store data about the object without being computers.

The devices that can be connected this way include pallets, totes, fixtures, jigs, tool holders, and even, when appropriate, workpieces. This technology pushes the edge of the systems, enables the collection of order of magnitude more error-free data and, once collected, to analyze it in a more flexible and more sophisticated fashion.

The suppliers of IIoT technology paint an attractive picture. According to GE publications, the Industrial Internet of Things (IIoT) combines machine-to-machine (M2M) communication, industrial big data analytics, cyber security, Human-Machine Interfaces (HMI) and Supervisory Control and Data Analysis (SCADA). Reading this list makes you wonder where the Internet in IIoT is, because none of the applications listed require it. Cybersecurity, in particular, is best achieved in a system that is entirely isolated from the internet! There is no physical way to hack into it and you first create a vulnerability by connecting it to the internet.

Vendors sell IoT devices that can be retrofitted or integrated into manufacturing processes. Figure 8.12 shows an example of DataMatrix readers implemented in a production line for plastic parts. Each part is marked with a DataMatrix code that is read by the readers (see the two black and yellow sensors) and provides an output of reader location, product ID, and a timestamp. This retrofitted setup allows the company full and near-real-time transparency into production status.



Figure 8.12 Two DataMatrix readers retrofitted into a production line of toilet actuator plates (Credit: Geberit)

8.4. Machine control

The software in the machine control layer of Figure 8.2 instructs machines to change settings, move workpieces, heat up or cool down a reaction chamber, and respond to anomalies reported by sensors. On new machines, the execution hardware and software is usually embedded in the machine and provided by the machine supplier, along with an off-line development environment to program the end-user's processes. End users often retrofit third-party controllers on older machines.

The reach of OT into an operation varies depending on its technology. In manual operations, it can provide task-specific operator instructions, possibly with step-by-step validation and prompting. In automated operations, programmability may be at the level of the smallest steps, as is the case in CNC machining or 3D printing.

8.4.1. Control of machine tools with CNC

Lathe, milling machines, grinders, and other machines to cut metals in a variety of ways have been around for 200 years. They are known collectively as "machine tools," and the technology used to control them today is the product of a long evolution. Originally, machinists controlled machines with cranks and mechanical dials, detected anomalies through their senses, and checked workpieces after an operation finished. Then came servomechanisms and analog dials in the 1940s, numerical controls (NC) in the 1950s, migrating to Computer Numerical Controls (CNC), later with digital readouts of machine parameters and settings.

There are two reasons to recall this history:

1. **It is still with us.** Machine shops accumulate machine tools over decades and multiple generations of control systems coexist on their shop floor, from the ancient to the state of the art.
2. **It is industry-specific.** CNC is for machine tools only. You need to know it if you work with machine tools but not with any other kind of production machinery.

CNCs use cryptic codes to describe steps. For example, the program statement "S1000 M03" means "turn spindle on clockwise at 1000 revolutions/minute." This industry also came up with a higher-level language called Automatically Programmed Tool (APT), which is more intuitive. In APT, the same statement is "SPINDL/1000, CLW."

An APT program has a "geometry definition" section in which a workpiece is described in terms of points, lines, planes, circles, and other geometric constructs that are then used to specify tool movements and cuts. Part drawings in electronic form in CAD systems contain this geometric data, and interface software exists to translate it into APT statements. A "postprocessor" translates the APT source into the codes for the machines to execute and the CNC's operator interface displays these codes as they are executing.

8.4.2. **Generic machine control with PLCs**

The Programmable Logic Controller (PLC) is the most common controller, but a few industries have their own special types. Machine tools have Computer Numerical Controllers (CNC), and semiconductor production machinery uses general-purpose computers as embedded controllers, while many simple machines, down to household appliances, use microcontrollers.

> **Definition**:
> A PLC is a computer that has been ruggedized and adapted for the control of manufacturing processes. PLCs turn power on and off, open and close valves in a prescribed sequence, and respond to sensor readings.

From their functional descriptions, microcontrollers can hardly be told apart from PLCs. Microcontrollers, unless explicitly stated, are not meant to survive for 20 years without stoppage in an industrial environment. PLCs are. PLCs have more Input/Output ports and they can put out a 240V signal. A signal from a microcontroller is going to be 3.5V to 5V, which means you need a separate switch to activate with this signal for high power.

PLCs can be supplemented by SCADA systems with human interfaces providing natural mappings to the equipment, as well as interfaces to plant-wide systems that match products to process programs and barcode or RFID-based auto-ID systems that validate product IDs for mistake-proofing.

Operators still trigger one action at a time, but, instead of being state changes on the machine, each action taken through the controller is a sequence of steps that transforms the product, thus making the operator interface centered on results rather than process. By pressing "Start," the operator starts a machine cycle during which, in principle, human presence by the machine is not necessary. However, many machines are subject to microstoppages – such as parts jamming in a chute – that the operator manually clears to restart the machine. As a result, the operator must remain close enough to the machine to be prompted by the stoppage and respond immediately.

The first programmable machine used in manufacturing was the Jacquard loom in 1804. It used a wide punch tape to weave patterns (see Figure 8.13) and is considered an ancestor of the computer.

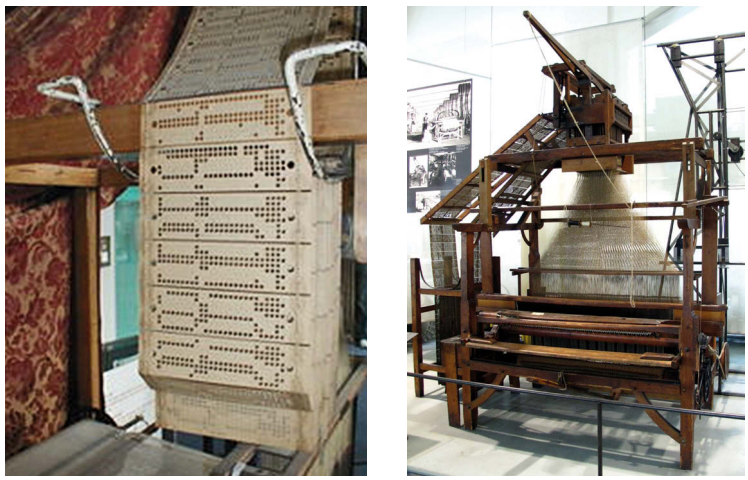


Figure 8.13 The punch tape of a Jacquard loom (Credit: public domain) and a loom (Credit: Dmm2va7 CC BY3.0)

PLCs are the most common type of programmable controller and are used in automated assembly and fabrication, and even in non-manufacturing machines like cranes. The services a PLC renders are shown in Figure 8.14. A PLC is generally speaking a ruggedized industrial computer.

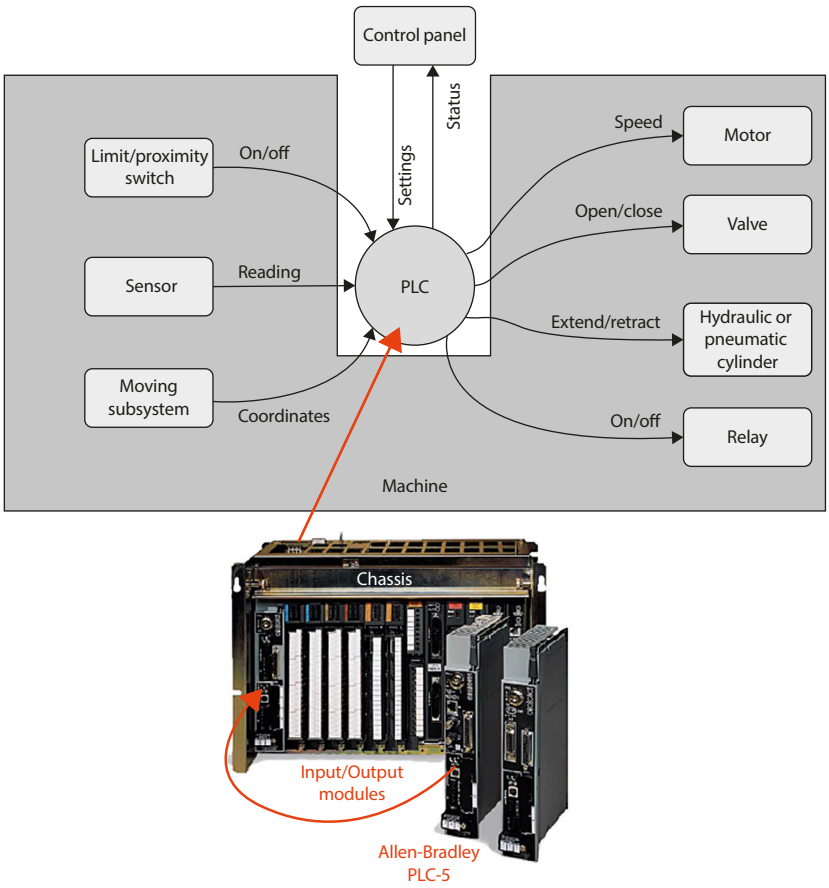


Figure 8.14 How a PLC looks and what it does

When a workpiece coming out of a chute triggers a switch, the PLC extends a hydraulic cylinder to push it; when the temperature in a kettle reaches a threshold, it opens a valve to dispense the content into another vessel; in response to the motion of a crane beam, it adjusts the power fed to motors, etc. And the signal from the PLC to the device does not only carry the information, but also the power to act on the device.

The PLC also reacts to operator actions on the control panel and presents a limited status display. When an operator pulls a stop rope, it is the PLC that stops the production line, sounds the alarm, and changes the values displayed on andon lights and boards. The PLC is programmed to execute a sequence of steps, where each step starts automatically as soon as the previous one has met its requirements.

PLCs are generic devices, used in all industries that have not felt the need to develop their own specific controllers. The machine tool industry uses CNCs instead, that are specialized in dealing with tool feeds and speeds, torque, and coolant flow. The semiconductor industry embeds powerful general-purpose computers as controllers into its machines.

PLC programs are generally written on a separate computer and then downloaded to the PLC. Except for providing the power needed to execute actions, the functions of a PLC can be emulated in software on a general-purpose computer. Many users, however, still prefer hardware PLCs for their reliability and device interface capabilities. The links shown in Figure 8.14 above are information flows. On current PLCs, there are no separate cables from the PLC to each device, and their communications occur through a network.

8.5. **Supervisory process control**

Operator interfaces have traditionally been the weak point of PLCs, and this has prompted the emergence of Supervisory Control and Data Acquisition (SCADA) systems as a remedy. As an acronym, SCADA has existed since the 1960s, and the definitions found in the literature make it sound almost indistinguishable from a PLC. In manufacturing today, however, SCADA systems are clearly distinct from and complementary to PLCs. Suppliers often combine SCADA with HMI for Human-Machine Interface as HMI/SCADA or SCADA-HMI. Figure 8.15 shows an example of a modern PLC from Siemens.



Figure 8.15 Modern PLC with HMI from Siemens (Credit: Siemens)

Modern PLC systems come in the form of software running on regular computers, ruggedized as needed to reside on the shop floor, communicating with one or more machines through their embedded controllers, most commonly PLCs (see Figure 8.16).

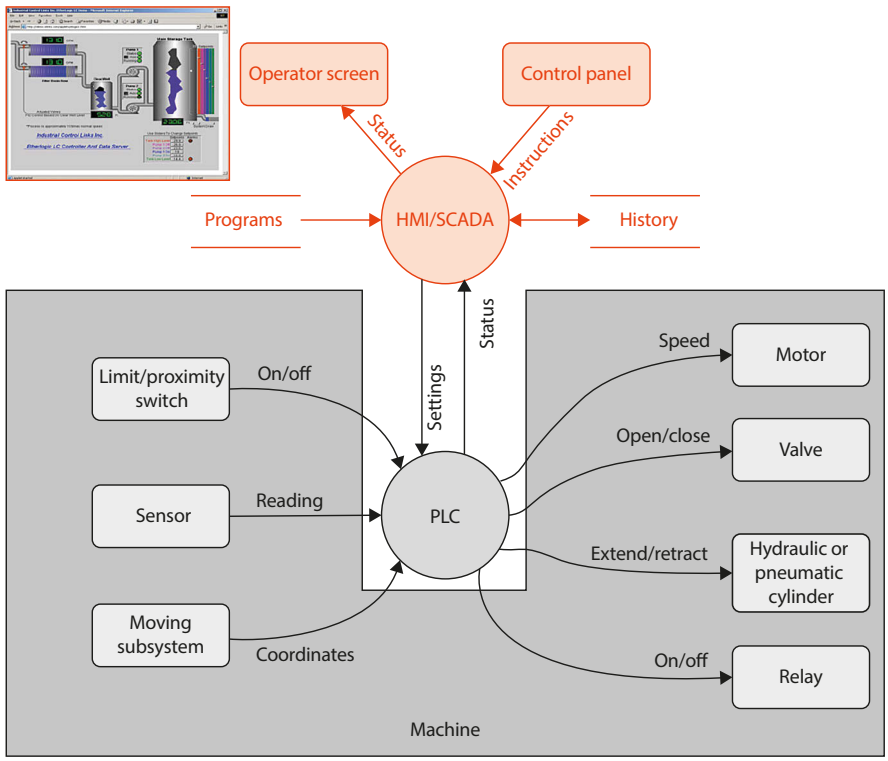


Figure 8.16 Enhancements of PLCs through HMI/SCADA

These systems are usually retrofitted to equipment by users and have the following additional functions:

1. Graphic editing and configuration for manufacturing engineers to design displays for operators, including icons for tanks, pipes, pumps, screw feeders, valves, and other devices, and the ability to link each screen icon with the sensors or actuators located on the actual devices, so that the SCADA screen shows the machine condition in real time, with the vital signs of each component displayed next to its image.
2. Storage and retrieval of historical values, and analysis based on more sophisticated models for alarm generation than just threshold crossing on a single variable.
3. Programming of machines. Programs developed off line with advanced tools can then be posted to different types of embedded controllers, without requiring the programmers to know the specific programming tools of each controller.
4. Storage, retrieval, and revision management of machine programs. The supervisory controller maintains a database of programs that are downloaded as needed to the embedded controllers of each machine. At this point, we have a control system that helps an operator attend to multiple machines but that does not hold or act upon any information about products. For example, it holds a local library of programs, but no cross-reference of programs to the products they are used for. The selection of the right program is up to the operator, based on information provided with the work pieces. Also, if there are multiple instances of this system in the plant, programs can be copied, but there is no mechanism in place to ensure that they are all at the same revision level, and they do in fact tend to diverge, causing inconsistencies in quality and productivity.

8.6. Operation support

There are many IT systems in the Operation support layer of Figure 8.2. Several of which are specialized software for specific operations, such as Warehouse Management Systems (WMS). In this section, we focus on the two most important IT systems for running manufacturing operations: MES and ERP.

8.6.1. Manufacturing Execution Systems (MES)

If the set of machines controlled by one single HMI/SCADA/PLC combination is unique, it is making only one product, and the process is stable, then a manufacturer does not need any further software to control and run that process. On the other hand, with low-volume, high-mix production, replicated production lines, or processes that are less than mature, this setup falls short. A Manufacturing Execution System (MES) can supply the missing data to the local controllers and keep them current. It can also feed machine, product, and process data back to other business software in real-time.

Figure 8.17 shows the control system of Figure 8.16, enhanced to receive the mapping of products to programs from an MES, and to use auto-ID technology to collect product item numbers from incoming work pieces.

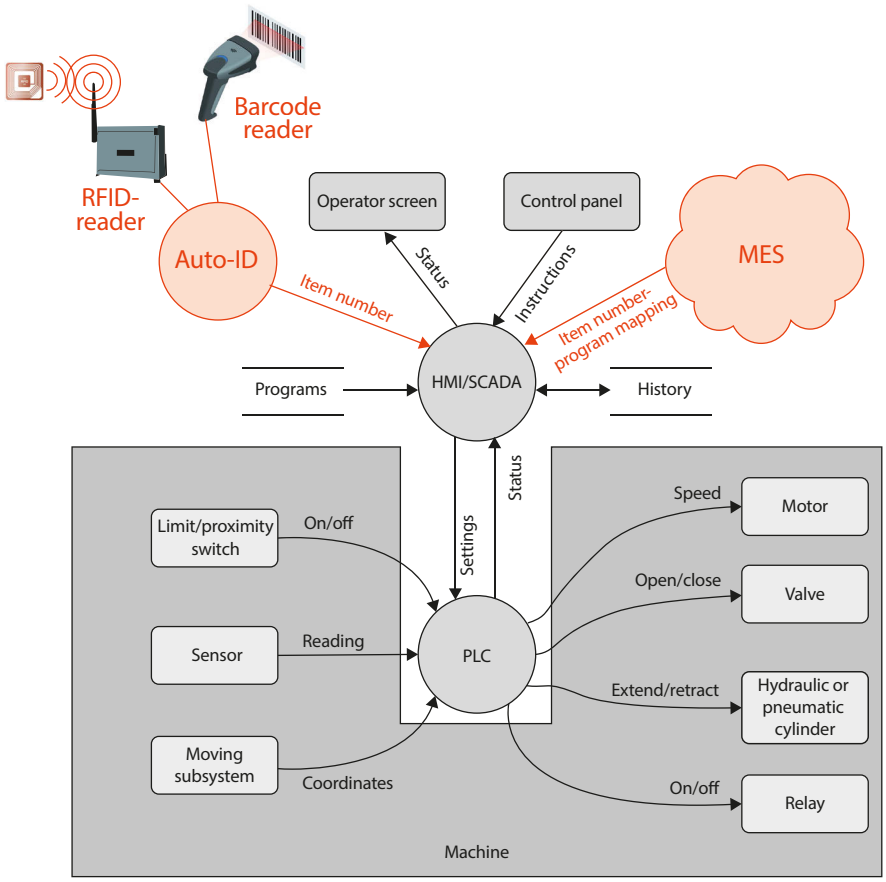


Figure 8.17 Enhancement of Figure 8.16 PLC through MES and Auto-ID

While it is not a complete picture of the functions of a machine control system, and the functional layers in any actual system may be different, it does show what can be expected in an advanced plant in 2006. For example, as indicated earlier, some industries do not use PLCs. The suppliers of the different types of subsystems, recognizing their limitations, have expanded their functions, blurring, for example, the boundary between PLCs and SCADA systems.

What makes MES distinctly different from other software is that it integrates business level data from the ERP system to the shop floor processes and helps manage, synchronize, and control real-time physical processes. An MES coordinates the execution of workorders with production scheduling and ERP systems. The extent of IT/OT integration and use vary greatly depending on needs and investments. MES applications also provide feedback on process performance and traceability.

However, an MES can have many other functions. Table 8.1 shows ten tasks listed by the Society of German Engineers (Verein Deutscher Ingenieure, or VDI) in the VDI 5600 standard.[1] Other standards exist too; all of which define and describe MES a bit differently – see for example the MESA-11 model from Manufacturing Enterprise Solutions Association International. Due to such differences, you may also encounter MES labeled as Manufacturing Operations Management (MOM).

Table 8.1 Ten tasks of a MES according to VDI 5600 (2016)

MES tasks	Description
Order management	Synchronize orders with production plans
Detailed scheduling and process control	Schedules production and controls execution
Equipment management	Provides a database of equipment
Materials management	Schedules materials to the processes
Human resource management	Schedules human labor to the processes
Data acquisition	Collects data from the processes
Performance analysis	Feedback on process performance
Quality management	Schedules quality control processes
Information management	Provides information to management
Energy management	Controls and manages energy consumption

Of particular note is the impact of the presence of an MES on the operator's interactions with the machines. The MES makes the machine controller aware of products. Before the machine starts processing a part, its product identification must be communicated to the controller, so that, with the help of the MES, it can retrieve the right program to execute.

MES is not a specific IT system but a class of different IT systems that try to achieve some of the tasks listed in Table 8.1. Today, although many companies still do not have a MES system, the relevance of a monolithic MES implementation is questioned. More and more tasks can be run via applications straddling local industrial IoT implementations and a central database. It is not a given that it is "better" to have a all-in-one system covering all the different tasks above.

1 VDI (2007). *Fertigungsmanagementsysteme: Manufacturing Execution Systems (MES)*. VDI 5600.

Stop and Think 8.4
Many manufacturers do not have a MES but work with workarounds in spreadsheet programs. Do they need one?

8.6.2. Enterprise Resource Planning (ERP)

Because the functional details of ERP is covered in Chapter 9, a short introduction suffices here.

Today, ERP broadly refers to commercial software packages that connects the firm's transactional processes. ERP systems have been the main applications supported by the IT departments. ERP is the latest of many generations of systems that, over 5 decades, have been successively called MRP, Closed-loop MRP, MRP-II, and ERP. They are increasingly ambitious iterations of a top-down approach that starts from planning functions at the top level and cascades down.

Integration or "business alignment" is is the core objective of ERP. It is however easier said than done. Users in the operational departments perceived the ERP models used for the senior managers to be simplistic and unable to serve their needs. ERP almost always spurs workarounds in the operational functions, in the form of specialized systems or internally generated Excel spreadsheets. Taken together, these systems and workarounds are data security hazards but they also create the towers of Babel, which is exactly what central IT wants to avoid.

The German IT company SAP and the US-based Gartner Inc. coined the term "Enterprise Resource Planning " (ERP) in 1990. It was a new name for what had became known as Manufacturing Resource Planning (MRPII) in the US and Produktionsplanungssystem in Germany. SAP, the leading player in the ERP marke, was established in 1972. Since ERP is a multi-billion dollar market, there are of course many other companies too; other large ERP vendors include Oracle, Microsoft, Infor, Epicor, IFS, and Sage. In addition there are hundreds of small- and medium-sized providers of generic and tailored ERP software.

Most companies have *one* ERP vendor, but it is almost never an integrated all-encompassing ERP system. Instead they can have up to several hundred IT systems that are partly connected, and partly disconnected to a core IT system (see Box 8.7). Even off-the-shelf ERP systems consist of bundles of software modules. Typical ERP modules are as follows:

- Finance and accounting
- Production
- Human resources
- Planning
- Inventory management
- Reporting and business intelligence
- Customer relationship management
- Sales and marketing

The greatest advantage of ERP is that is seeks to establish a single-point-of truth. Therefore master data management (covered in Chapter 10) is essential for its performance. The quality of an ERP system relates directly to the quality of the master data it uses for calculations.

Box 8.7 **IT's a jungle out there**

Companies' IT landscapes often differ radically. While many companies use the same ERP vendors such as SAP or Microsoft Dynamics, data storage and analytic services like Amazon Web Services, and common office software, the amount of different IT systems in almost any manufacturing company is usually staggering. For a mid-sized factory, it is not unusual to find 200 or more different types of IT systems.

What goes by the name of "ERP" is in most cases a collection of numerous different software applications, possibly sharing a database. Mapping a company's IT infrastructure is therefore moving through a jungle (Figure 8.18). If mapping it is a challenge, imagine the difficulty of changing it!



Figure 8.18 Trying to map a company's IT systems is akin to moving in a jungle (Credit: Dustin Luther/Twenty20)

However, there are great differences between industries. In a semiconductor plant, all machines typically belong to the same technology wave and have similar levels of control technology. In a machine shop, on the other hand, you have machines of multiple vintages accumulated over decades, ranging from analog devices like copy lathes to transfer lines and Flexible Manufacturing Systems comprised of CNC machining centers connected by a shuttle system and with a central controller routing parts between the machining centers, possibly supplemented by robots.

Manufacturing, as an industry, has never been at the leading edge of information technology. The oldest clay tablets we have from Mesopotamia contain shop keepers' books, not metallurgical methods. Today, the most advanced IT is deployed first in financial services, defense, or gaming, years before it finds its way onto factory floors. For this reason, to plan for the future, manufacturing professionals need to stay aware of developments in IT in other industries.

8.6.3. **Software-as-a-service apps**

The latest development in the manufacturing IT space is platform-based apps delivered as software-as-a-service (SAAS). Different from ERP's goal of being an integrated, all-encompassing business system, software vendors are offering apps or app databases as the *system of interaction*. These apps do not care where data comes from as long as it can be read or written. Instead, they seek to offer tailored solutions to manufacturers with intuitive user interfaces.

One example is Tulip that offers a "no-code" app platform that can be used by frontline workers. Their apps vary from simple checklist apps to MES-like functionalities. Such apps can draw data from any OT/IT system, including IIoT technologies. The advantages are numerous compared to traditional IT system implementations:

- **Near real-time control.** Apps can update instantly, which is often not the case for MES systems.
- **More intuitive interface.** Can be used by all.
- **Simpler start.** Anyone can start and test an app; no need to implement a full system.
- **Easier to scale.** Apps in the database can be reused elsewhere.
- **Easier to upgrade capabilities.** Modular build up makes it easier to integrate the latest developments.
- **Easier to develop applications.** The modular architecture makes new app development simpler. Can be done locally with production engineers and IT staff. Often offered with "no code" or "low code" user interfaces.
- **Easier to use.** Empowers operators and engineers to do their jobs with better information.

Due partly to opportunities offered in platform-based SAAS, an increasing number of startups are entering the manufacturing OT/IT domain. Scaling new solutions have always been difficult in manufacturing, but the new platform-based app warehouses enable faster trial and error for manufacturers without the need for monolitic implementation of large IT systems.



Figure 8.19 Apps are slowly making their ways into factories

8.7. Engineering and management support

The point of enterprise IT and OT is to support operations, assist in problem solving, inform planning decisions, accelerate product development, and new product introduction. It is also supposed to help the organization learn and accumulate knowledge. In later chapters, this book discusses IT systems that support production *improvement* activities; the following sections discuss the most usual business software that supports engineers and managers: Spreadsheets, slideware, and collaborative software.

As of the early 2020s, however, these systems are mostly used as sources of raw data for individual engineers and managers to post-process and communicate on with their own tools, the most common being electronic spreadsheets, slideware, and email. They are so ubiquitous that few see their limitations and the need for more powerful ones, including state-of-the-art analytics software, writing tools that produce documents with more than bullet points, and collaboration tools that simplify team communications and knowledge management.

8.7.1. The limits of spreadsheet workarounds

The original appeal of electronic spreadsheets is that they emulated paper spreadsheets (see Figure 8.20) that provided a grid on which you could draw tables of numbers and lists of parameters. It eased the way into the computer age for number crunchers but it was a horseless carriage. The electronic spreadsheets, from Visicalc to Lotus 1-2-3 and Excel, served as an introduction to IT for several generations of number crunchers, most of whom never looked further.

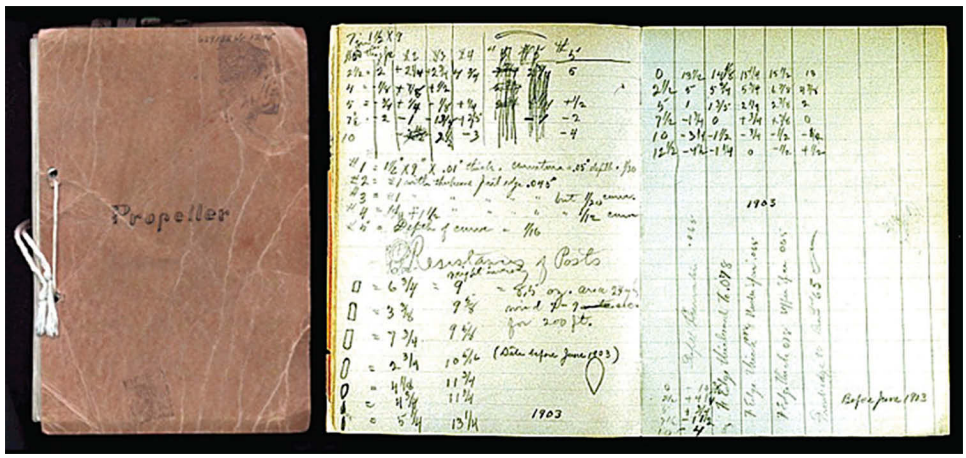


Figure 8.20 Wright Brothers' spreadsheet from 1901

This included not only bookkeepers and accountants with paper ledgers but also managers who used pocket calculators to add up columns of numbers from printed reports out of corporate mainframe systems, and engineers who retrieved function values from handbooks of mathematical tables.

For all these constituencies, the electronic spreadsheet was faster, less error-prone, and more powerful than their existing methods. To serve them, electronic spreadsheet suppliers have kept adding features, and discounting academic licenses to make sure that the next generations would be addicted to their products.

The way data is stored and retrieved should be tailored to the uses made of it. It is currently not the case in most factories. ERP, MES, and other systems are used for *transaction* processing, which only involves lookup. For search and analysis purposes, engineers and managers extract tables from these systems and process them in electronic spreadsheets. Figure 8.21 shows how data is organized on a spreadsheet.

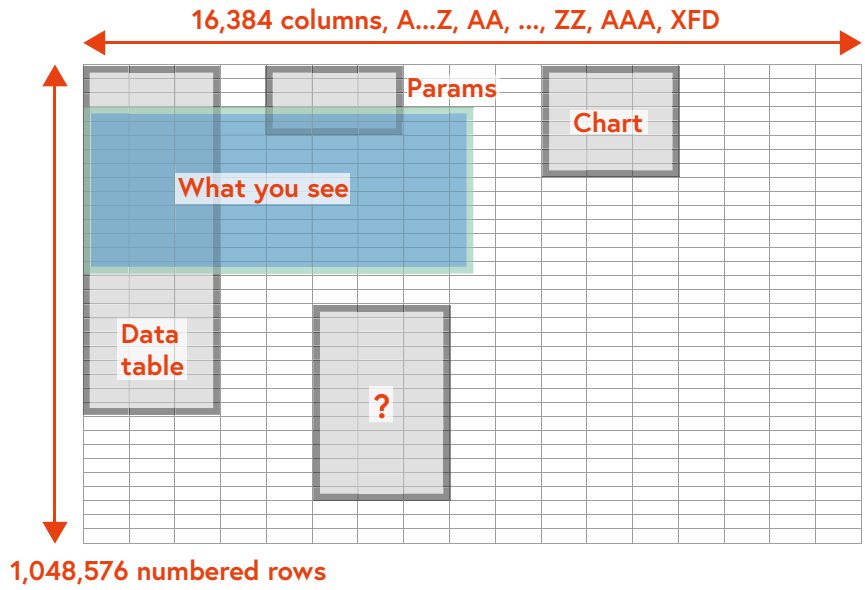


Figure 8.21 Spreadsheet structure

The grid of cells, today, is unnecessary. It complicates the work and makes it error-prone. Work with tables and lists is best done with tools that directly operate on tables or lists.

Spreadsheets proliferated in factories because they provided immediate, short-term solutions. The master data copied into them, however, was not updated automatically. When working with others, individuals passed around copies of their workbooks to colleagues, who modified them without changing the file names, and, in turn, passed them to others.

When individuals change jobs, the spreadsheets they developed for their current job are not commonly used by their successors, who don't understand and don't trust their predecessors' formulas. They often have good reasons, as these spreadsheets are riddled with errors and inconsistencies. When individuals leave the company, you practically cannot prevent them taking copies of their workbooks with them, with proprietary information. It is a data security vulnerability.

This is no reflection on electronic spreadsheet products. They are useful tools for an individual to run calculations and plot the results but they were never intended to serve as a shared data repository for an organization of 1,000 members. It came to be used this way because it sort of worked in the short term and it was the only tool engineers and managers knew, often as a result of having used student versions in school. Over decades, it became, at work, a standard tool given to every professional employee, with strong encouragement to avoid using any other.

8.7.2. **The limits of slideware**

Slideware is software to generate and project visual aids on a screen for oral presentations. Since the 1990s, however, it has been used as a general-purpose tool to generate documents ranging from strategic plans to business proposals and project reports that are intended to be read standalone rather than presented.

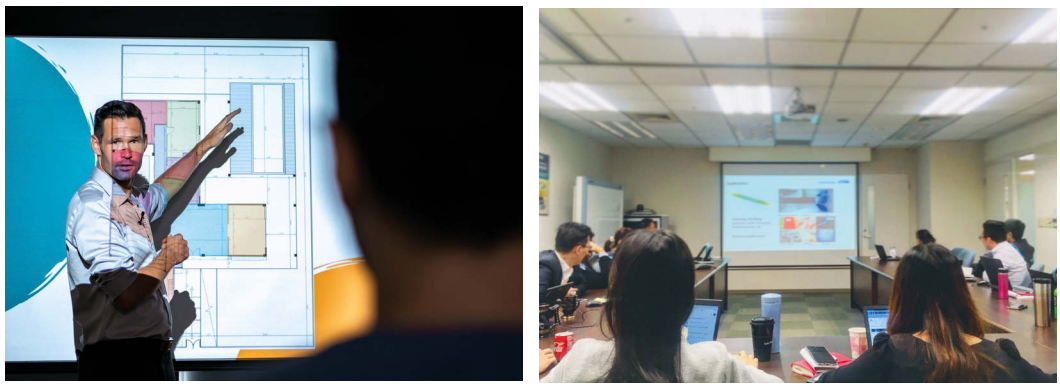


Figure 8.22 Slideware is everywhere in corporate communication

The advice books on presenting are all about selling, inspiring, and entertaining. Slides with bullet points, graphics and animations gloss over problems, making ideas appear sounder than they really are. This works for selling but not for project management or problem solving. Every day, millions of slides are shown by manufacturing people reporting on the status of a project or the results of the last quarter, for the purpose of starting a discussion, training, solving a problem, or laying out options. Others are used by teachers, instructors, and coaches to transfer knowledge and skills, with a focus on understanding, application, and retention.

If you work in the media, or in the marketing or public relations department of a company in another industry, the medium is a major part of the message, and communications is a profession. On the other hand, in manufacturing operations, engineering, or R&D, the message is the message, and communication is a set of skills everybody needs. Business communication is not a minor topic, and neither is the technology that supports it.

What can an organization do to stay current? Beyond technology watch on the latest tools available, new hires are a valuable source of information, particularly when they have several years of experience with tools you don't know. Before indoctrinating them in the exclusive use your standard kit, you might want to debrief them on possible ways to enrich it.

In every organizations, there are individuals who are frustrated with the standard kit and secretly use tools the IT department does not support. Instead of being disciplined for this, they should be encouraged to experiment in an organized manner. A team, for example, may be allowed to use a different project management package, but required to report on it to support a decision to adopt it for the whole organization, or not.

8.7.3. **Team communication with email and beyond**

In the 2020s, modern collaboration tools are available from many sources at low cost. Most manufacturing companies, however, still rely on email. Email, today, is the primary medium of person-to-person business communication across organizations and with strangers. It is used to send meeting agendas and minutes, submit proposals, send invoices, receipts, reminders, complaints, etc. It is not, however, the best medium within an organization. Email is insecure and mistake-prone, in many ways that do not even require an attack from a hacker. That email places copies of messages and attachments in each recipient's mailbox, which makes protection of proprietary information nearly impossible.

In addition, the most commonly used email clients send messages without prompting the user for confirmation. They send messages to the wrong list of recipients in just one click, after which the sender cannot recall them. Almost every user of email has experienced sender's remorse but been unable to do anything to correct such mistakes. They go into the sender's permanent record.

Forums and *discussion groups* have been the most successful alternative to email for team communications. Team participants post messages to spaces shared with a managed team roster, and upload documents to a shared library instead of sending them as attachments. The technology is as old as email and is the foundation of social networks. Many manufacturers take advantage of the availability of smartphones and use chat apps, like Telegram, WhatsApp, etc., to communicate in teams (Figure 8.23).

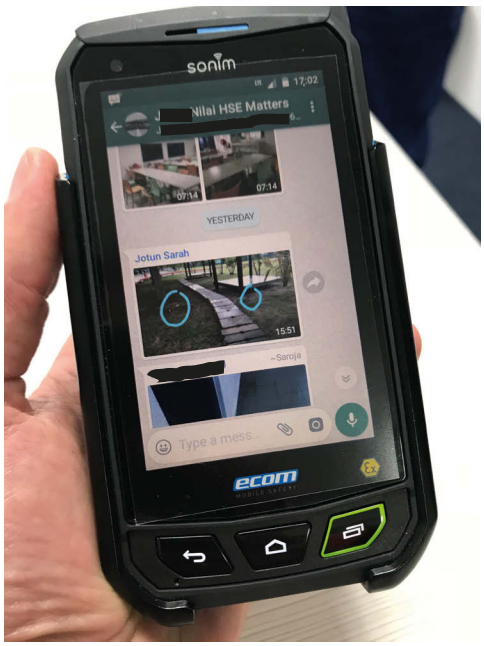


Figure 8.23 Some companies use chat apps to efficiently communicate in teams

In the 1990s, after failing to convince consultants to share their expertise in its enterprise "knowledge management" system, a major consulting firm adopted a tool with messaging and document sharing capabilities. There, consultants asked for help on forums; their colleagues actually *responded*, and shared their experience. Today, software users who are stuck turn to a question-and-answer site called Stackoverflow operating along the same lines, with millions of users.

8.8. Management of manufacturing IT/OT

Many manufacturing companies struggle with the management of IT/OT, which they perceive as a distraction from their mission of *making things*. IT/OT is not only a substantial investment in hardware and software, implementation, training, but also an ongoing operational expense. As such, it needs a home in the organization structure.

8.8.1. The IT department versus Engineering department

We have been using "IT/OT" as an umbrella term for all the software applications that support the activity of a manufacturing organization. Most manufacturing organizations have an IT department and entrust OT to the engineering group in charge of each operation. The IT department provides the hardware and the applications that interact with human users and with other applications; various engineering groups, the controllers that take input from sensors and issue output to actuators. For example:

1. Telling an AGV to move from point A to point B is an OT function. A production control app identifies this move as necessary and gives the order to do it. This is IT. In response, the computer embedded in the AGV plots a path from A to B and drives the AGV while dodging fixed and moving obstacles. This is OT.
2. Failure analysis on a product unit after final test is an IT function. Making a machine perform a sequence of steps while keeping its vital signs within specified limits and taking in-situ measurements on work pieces is OT.

While the line between IT and OT is blurry, there are several reasons to draw it:

1. The stakes in getting the systems to work right are different. In IT, errors may lead to bad business decisions; in control, they may destroy products and equipment, and injure people.
2. The technical skills required for IT and OT are different.
3. As long as IT and OT are managed by different departments, their purviews need to be clearly stated.

The bundle marketed as Industry 4.0 in Germany spans both IT and OT. So do similar initiatives like *Smart Manufacturing* in the US, *Made in China 2025*, and *Society 5.0* in Japan.

The key reason given for the applications to be managed centrally by the IT department is that they share the infrastructure of networks, servers, workstations of different types, and a set of approved and supported software tools. To a large extent, however, this organization structure is a legacy of the mainframe age, when all the data processing was done in a single computer for the whole plant, interacting with dumb terminals keystroke by keystroke.

Innovation since the mainframe era has technically enabled the various departments to go their own way and acquire their own systems, an evolution that IT departments have been resisting, leading to the paradox that employees of major corporations often have more advanced tools in their private lives than at work.

In a 1947 lecture, Alan Turing had warned about the response to continuing innovation by an IT profession that didn't yet exist and could not be offended by harsh words.[2] He said:

> *"They may be unwilling to let their jobs be stolen from them in this way. In that case they would surround the whole of their work with mystery and make excuses, couched in well chosen gibberish, whenever any dangerous suggestions were made."*

More than 70 years later, engineers and managers in operations functions may recognize in Turing's words their interactions with corporate IT. When they try to innovate with data, they may experience limited help from the IT department. From a *corporate* perspective, IT has a point in trying to prevent a free-for-all service. If every department goes its own way in IT, the factory and the company become a tower of Babel, where data generated in multiple systems cannot be cross-referenced. As of the early 2020s, however, IT departments have been unsuccessful in preventing this outcome. The following sections point to ways of doing this successfully.

The effective programming of production machines requires the complementary skills of programmers, manufacturing engineers, and experienced operators. The required collaboration has several prerequisites:

- Management must be aware of the need. Many manufacturing and engineering managers are still new to the challenges of developing any kind of software in house and sometimes do not even consider that NC or PLC programs are software, just like the packages used to run accounting. As a consequence, they tend to underestimate the skills required.
- Advanced tools must be provided. Development environments supporting sequential function charts for PLCs, or CAD interfaces that partially automate the generation of APT programs speed up program generation, make it less error-prone, and make programming easier for domain experts to learn.

The introduction of sound development practices and advanced tools is rarely welcomed by programmers who have spent months or years mastering the lower-level techniques. Likewise, in society at large, attempts to simplify spelling are most opposed by those who have taken the trouble to master the old, complex rules. They now must forget and learn new ones, which makes it easier for others to acquire a skill that had previously been special. Advanced, high-level program development tools enable more people to get involved with this activity faster, which of course is in the best interest of the organization, but management must see to it that they are introduced in a way that involves the existing programmers.

Because process programs embody process engineering know-how and company trade secrets, their value as an asset far exceeds the cost of generating them. The key points to manage them accordingly are as follows:

1. Develop your process programs in-house by the best qualified engineers possible. Do not leave the generation of these programs to operators and machinists alone.
2. Keep your process programs in a secure repository, and download them to the machines through a network, rather than on physical media that can all too easily leave the plant.
3. Design process programs so that product-specific logic is insulated from machine-specific commands that need to be remapped to transfer the program to a different machine. For long-life

2 Turing, A. (1947, February 20). *Lecture to the London Mathematical Society*. URL: https://www.vordenker.de/downloads/turing-vorlesung.pdf

products, well-engineered process programs can outlive the machines they were written for.

In the 2020s, restricting the hardware, system software, and apps used in a manufacturing organization to a standard set defined and maintained by IT/OT is neither necessary nor sufficient to ensure that the participants communicate. It is ineffective.

It doesn't matter whether individual computers run on Windows, Linux, OSX, iOS, or Android, anymore than it matters which brand of ballpoint pen engineers use. What matters is that the devices should all be able to exchange messages, read from and write to repositories of shared data, while protecting proprietary information.

It is a *human* challenge, not a technical challenge. Individuals who share data – like production operators and materials handlers – must use the same formats, relationships, and versions. Departments that share data with other departments must do the same for what they share, etc. Departments have reasons to group products in different ways. For Engineering, it may be by process and feature similarity; for Production, by volume and stability; in Sales, by market segment; in Accounting, by revenue or profit.

The different groupings, however, must be properly cross-referenced for the whole organization, so that you can retrieve the data about products made with a process P, at volume ≥V, for 16 to 24-year old females, with a value added ≥25%. The IT/OT of most manufacturing plants today is incapable of producing this kind of information on the fly.

8.8.2. **A bottom-up approach**

A common, mistaken belief is that there is no point in trying to improve the performance of legacy systems. A new all-in-one that will be implemented two years from now will solve all the problems, and the supplier's consultants will work with you to reengineer your business model and fit the system's capabilities. This is the next iteration of the top-down strategy and not likely to deliver on its promises.

Instead, the whole range of IT/OT and process control should be the object of continuous improvement. Many of the "Industry 4.0 enablers" can be retrofitted to legacy systems, yielding benefits in months rather than years, at a much lower cost than replacing the legacy systems. In fact, some of the enablers, like data warehouses, are specifically intended to get value out of the data in legacy systems. In addition, as in production, this continuous improvement effort develops skills that make the organization a savvy buyer when actually replacing the legacy systems or designing systems for new plants/lines.

As for production lines, part of the talent needed to pursue continuous improvement in IT/OT and process control is already present in most manufacturing organizations, in employees who like to tinker with software. Left on their own, they are the ones producing the problematic Excel spreadsheets discussed above; organized in a coordinated approach, they can help improve the business. Additional talent is needed to provide the tinkerers with better tools, train them, and organize their efforts so that their contributions coalesce into a coherent system. Recruiting this talent may be the most difficult challenge.

Summary

The most essentials points covered in this chapter were:

1. Modern manufacturers are dependent on their IT/OT systems. We separate between OT and IT, although with no clear border. OT is Operational Technology and deals with machine action. IT is Information Technology and deals with information to humans and other IT systems.
2. An IT/OT stack helps understand the logic and landscape of IT/OT systems, but the real world is more messy.
3. The PLC is the workhorse of automation.
4. IoT allows near real-time collection of data from the shop floor.
5. SCADA systems integrate distributed PLCs and enable coordinated control of machines on the shop floor.
6. The EDI standard enables automatic transfer of transaction data in supply chains.
7. MES connects shop floor data with business transaction data.
8. ERP is a collective term for software that capture business transaction.
9. The most used software in manufacturing is MS Excel.
10. The lines between IT and OT are blurring and new technologies can connect shop floor machines directly with a cloud, which makes the management and ownership of IT/OT a real challenge to solve.

Test yourself

1. Why does the difference between OT and IT matter?
2. What is the difference between a data warehouse and a data lake? When would you recommend the use of each of them?
3. Explain how a PLC work.
4. What are XML, YAML, and JSON examples of? Why do they matter?
5. What is the purpose of MES and what does it do?
6. What is the purpose of ERP and what does it do?
7. What are pros and cons of IoT?
8. What are the key challenges relating to data transfer across supply chains?
9. Why is MS Excel the most used tool in manufacturing? Is it the best?
10. What department if any should lead a manufacturer's digitalization efforts?

Further reading

Bauernhansl, T., Ten Hompel, M., & Vogel-Heuser, B. (Eds.). (2014). *Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung· Technologien· Migration*. Springer.

Efron, B. & Hastie, T. (2016). *Computer age statistical inference: Algorithms, evidence, and data science*. Cambridge University Press.

Kagermann, H., Wahlster, W., & Helbig, J. (2013). *Recommendations for implementing the strategic initiative Industrie 4.0*. Acatech. National Academy of Science and Technology, Germany.

McCrady, S. G. (2013). *Designing SCADA application software: a practical approach*. Elsevier.

Meyer, H. (2009). *Manufacturing execution systems: optimal design, planning, and deployment*. McGraw-Hill Education.

Rehg, J. A. (1995). *Computer-integrated manufacturing*. Prentice-Hall, Inc.

Rinaldi, J. (2016). *OPC UA-Unified Architecture: The everyman's guide to the most important Information Technology in industrial automatio*n. CreateSpace Independent Publishing Platform (Amazon).

Schwab, K. (2016). *The fourth industrial revolution*. World Economic Forum.

Thames, L. & Schaefer, D. (Eds.). (2017). *Cybersecurity for Industry 4.0*. Springer.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

9 Planning and Control

The key role of manufacturing management is to plan activities for the coming days, months, and years, schedule shop floor events for today, and control the execution of schedules and plans. Without plans, we have no direction on what to make, when, or where. We should, however, be ready for markets, politics, natural disasters, or industrial accidents to force us to pivot.

Plans need to be turned into schedules, which detail the exact timing for when activities will be carried out in the *future*. As the manufacturing organization executes plans and schedules, management needs to detect and respond to any discrepancies that arise. The control functions report progress and performance back to the planners, and the key is the kind of countermeasures they can apply to get back on schedule or adjust expectations of output.

In factories, practitioners experience regularly that the "plan is wrong" and develop ways to work around it. Depending on how a manufacturing organization designs, lays out, and allocates products to production lines, the tasks of planning, scheduling, and control can be either simple and straightforward or mind boggling in complexity. This chapter will help you navigate and overcome these challenges.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Understand the role of planning, scheduling, and control in manufacturing.
2. Know how to generate a Master Production Schedule (MPS) from a sales forecast.
3. Turn MPS into production schedules and plans.
4. Implement a balanced workload (heijunka).
5. Implement autonomous shop floor control systems such as two-bin systems.
6. Explain how the execution of the plans on the shop floor is followed-up and controlled.
7. Make use of shop floor control techniques based on cards or tokens such as different kanban systems.
8. Make use of visual techniques like andon to control running operations.
9. Use Materials Requirements Planning (MRP) to support production planning.

DOI:10.4324/9781351110310-12

9.1. The planning challenge

Today, almost all planning, scheduling, and control tasks are executed with the help of IT, but there is more to it than IT. Scheduling, in particular, is where the world of data engages with the world of materials. It entails issuing instructions for people and machines to act on materials, and doing it effectively frequently requires supplementing IT with autonomous low-tech solutions on the shop floor.

Production planning, scheduling, and control is a challenge that everyone understands and thinks they can meet with common sense. If we assume that the bills of materials, routing, and process times are known, then we have all we need to make the plan when orders with item numbers, quantities, and delivery dates come in. Yet, for this "simple" task, researchers and practitioners have suggested hundreds of different methods.

9.1.1. Planning horizons

It is common to split between short-, medium-, and long-term planning horizons, but their meaning varies with each industry's "clockspeed."[1] Clockspeed, in this context, is the rate of change in process technologies and new product introductions.

The planning horizon won't be the same if you are making commercial airplanes to orders placed years in advance or toys for next Christmas. A common pattern is for short-term planning to prepare for the coming weeks, medium-term for the next 6 months, and long-term for the next 3 to 5 years.

The nature of the planning decisions and options available vary across these horizons. Short-term planning, for example, may involve asking employees to work overtime or to take time off; medium-term planning, opening a third shift or overhauling a major piece of equipment; long-term planning, building a new factory or reallocating products among factories.

Table 9.1 Examples of high and low clockspeed industries

High clockspeed industries	Low clockspeed industries
• Consumer electronics such as computers, smartphones, gaming technologies • Semiconductor industry • Toys • White goods	• Aircraft • Building materials • Machinery • Food and beverage • Metals • Pulp and paper

With high clockspeed, managers experience the consequences of their decisions first-hand; with low clockspeed, they may be retired or dead before these consequences play out. Low clockspeeds challenge managers to worry about what will happen, say, 30 years from now.

1 Fine, C. (1998). *Clockspeed: Winning industry control in the age of temporary advantage.* Perseus.

Box 9.1 **Planning horizons in the metering industry**

For decades, the US electricity meter market was stable. A meter would last 20 years and a manufacturer could serve a market with a stable installed base of, for example, 20 million meters by making 1 million units/year every year (Figure 9.1 left). The demand was easy to forecast, until the market was disrupted by smart meters, communicating with the electricity company over the power lines (Figure 9.1 right).



Figure 9.1 Traditional electromechanical electricity meter (left) and modern digital meter (right)

The disruption started in the early 1980s, when electro-mechanical meters were first able to respond to signals sent over the power lines, to switch between day and night rates at fixed times. Soon, consumers were starting their washing machines as soon as the night rate kicked in, causing surges in consumption. The utility companies responded by staggering the switch over one hour to smooth the surge, which required multiple messages to be sent and the ability for meters to respond to one only. These meters still required human readers. Today's meters are fully electronic and *send* meter readings over power lines, and their installation is opposed by some users for privacy concerns.

The conversion of a 20-million-meter market to this new technology in 5 years translates to a temporary demand of 4 million units/year, with new production processes. Once this market is saturated with new meters, its demand will drop to a much lower level than the replacement of traditional meters. To survive then, the manufacturer must either diversify its product line or export. Either way, planning is no longer trivial.

The electricity meter market, however, still has a longer planning horizon than, say, electronics assembly. The majority of the products made at any time have been introduced within the past 12 months and much of the assembly work is done by *contract manufacturers* for products designed by other companies and sold under these other companies' brands. The workload of a contract manufacturer is a mix of products, from small and simple to large and complex, and from low volumes to high volumes. It includes new products that may be duds or runaway hits. Planning for the next month by product is feasible based on current activity, but for the next year only for capacity on aggregate volumes.

9.1.2. **Responding to demand variability over time**

Figure 9.2 illustrates two strategies for dealing with seasonal variations. If we know what item the peak is for and can anticipate it, then we can build inventory ahead. This is usual in many industries such as the seasonal sports industry (Figure 9.3 shows snowboards). Otherwise, we have to secure the capacity to meet the peak demand.

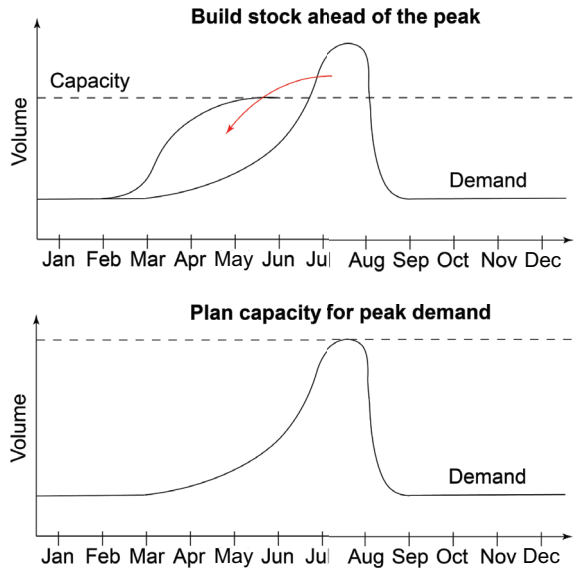


Figure 9.2 Responding to seasonal variations

The two approaches can be combined, for example by building ahead standard components requiring expensive equipment, and plan for enough manual assembly lines to assemble custom products to order as various configurations of the standard components.

Securing the capacity to meet peak demand does not necessarily mean building it in-house. Another approach is to make an agreement with a "coopetitor" to become a second source and provide surge capacity. This is practiced for example among suppliers of parts to machine makers, a business in which demand may double overnight. It avoids overinvesting, at the cost of foregoing part of the income generated by the peak demand, and at the risk of nurturing a competitor.



Figure 9.3 Snowboard production is subject to high seasonal demand met by building inventory in advance

9.2. **Production planning overview**

Figure 9.4 shows the typical information flows in production planning. It is drawn with data flow diagramming to provide a high-level, synthetic view of a complex system, strictly in terms of who receives what data from whom. By omitting many technical details, this style of diagram makes it possible to show a complex system in a single sheet. It is not a complete picture, particularly because it does not include *exception handling* that accounts for much of the information exchange between the focal firm and its customers and suppliers. The diagram omits the timing and the sequencing of the data exchanges. It also does not show planned versus actual reporting or supplier performance monitoring. It only represents a common pattern in manufacturing. The central information flows in Figure 9.4 is similar to what is done in ERP.

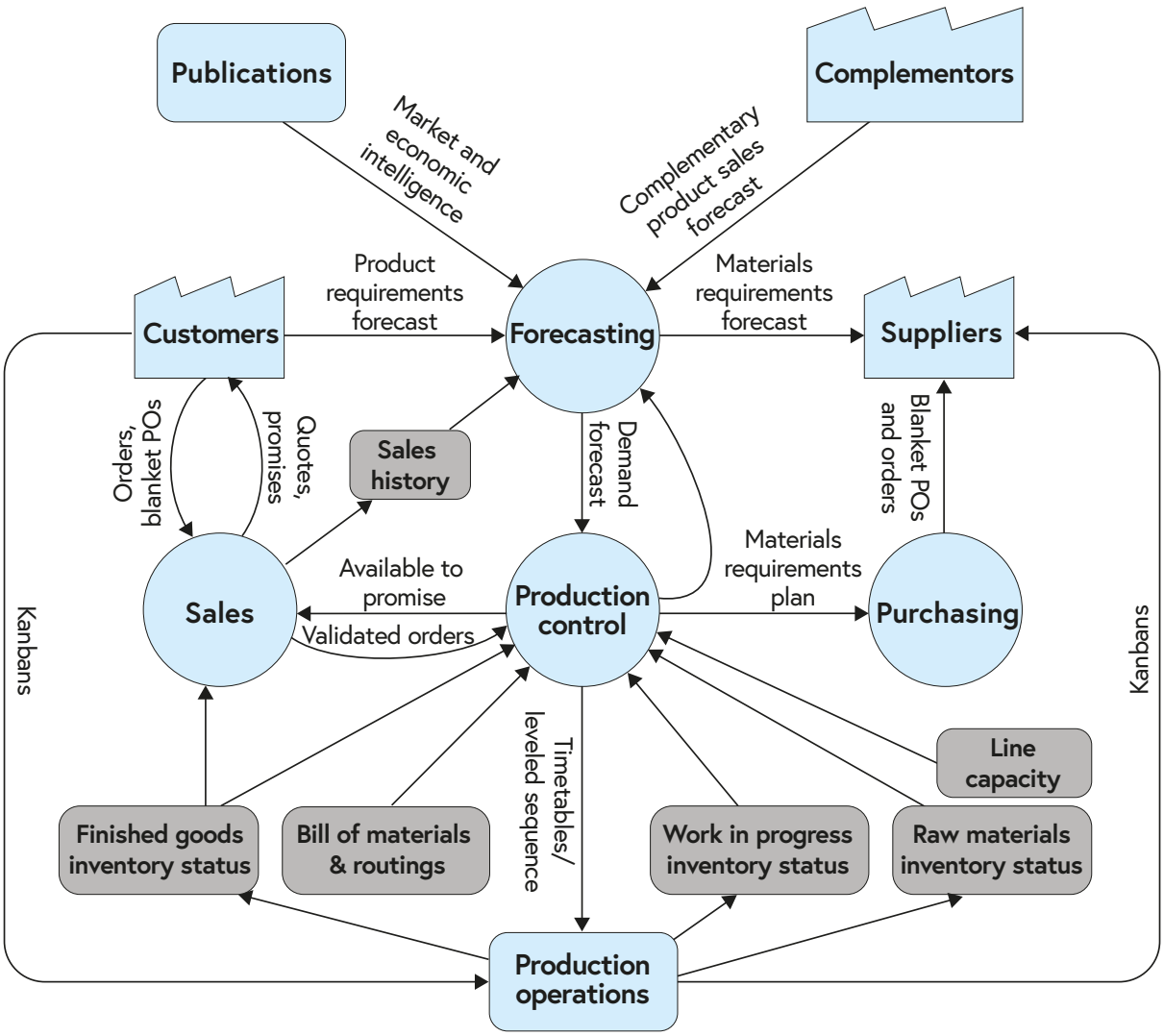


Figure 9.4 Planning and forecasting information flows

The key elements of the diagram are as follows:

- The four *round* bubbles – Forecasting, Sales, Production Control, and Purchasing – are firm-internal support activities. They are detailed below.
- The two *sawtooth factory symbols* in left and right for Forecasting represent customers and suppliers.
- The blue *rectangle* Publication designates external sources of market and economic intelligence. It includes industry forecasts in trade publications, general economic forecasts, monetary policy decisions by the central bank, war and peace, and events like earthquakes, floods or hurricanes that may affect buyer behavior in the near future.
- The *factory symbol for* Complementors designates another external source of input to Forecasting. "Complementor" is a term coined by Adam Brandenburger and Barry Nalebuff[2] to designate companies serving your customers with products that enhance the value of yours. Makers of carrying cases, for example, are complementors to smartphone makers; oil companies, to car makers; software companies, to computer manufacturers.
- The *arrows* show data flows, and are all marked with labels. Formats and means of transmission are omitted to keep the diagram focused on meaning. At this level, what matters is that it's an order or a quote, not whether it's a piece of paper or an electronic message in XML.
- The *grey rectangles* are stores of data needed for planning and scheduling:
  - Finished goods inventory status: How many do we already have on stock?
  - Bills of materials and routings: What parts go into the product and what capacity do we need to make them?
  - Work in progress inventory status: How many are we already making?
  - Raw material inventory status: How many can we make with the parts we hold?
  - Line capacity: How many can we make per period with the available capacity?
- The arrows marked with "*kanban*" flowing from Customers and to Suppliers are orders that do not need to transit through Sales or Purchasing because they occur within agreed frameworks set in blanket purchase orders. They are triggered and limited by consumption.
- Ultimately, *Production Operations* attempts to execute production schedules and returns status data.

Forecasting

The complexity of the actual processes is hidden in Figure 9.4. "Forecasting," for example, covers both calculations and human judgment. While sales history data can be fed to an algorithm, the other inputs are more complex. Customers' product requirement forecasts, for example, are issued by people working for other organizations and with varying skill levels. One dealer may systematically overestimate future demand, while another has the opposite bias and routinely exceeds expectations.

Production control

Production Control is responsible for planning, scheduling, and controlling shop floor activities. It is done by production planners and production/line/area managers and goes under different labels in different firms. For example, Production Planning is also common. After being passed on to Production Control, the demand forecast is exploded backwards in time based on component quantities in *bills of materials* (BOM) and *routings*, *production capacity*, and *inventory levels*. BOMs, routings and capacity are

2 Brandenburger, A. & Nalebuff, B. (1996). *Co-opetition*. Currency/Doubleday.

master data, discussed in Chapter 10. The resulting information is shown in Figure 9.4 is as used in two ways:

1. To issue materials requirements forecasts to suppliers
2. To size blanket purchase orders that provide the commercial framework for ordering from suppliers through kanbans.

In Figure 9.4, this is shown as Forecasting and Purchasing processes using the same output from Production Control. Both of these communications are usually handled by Purchasing, but we separate them to emphasize their differences, particularly in terms of responsibilities with respect to the suppliers. While a blanket purchase order simply records agreed terms and conditions, a forecast is an input to business decisions that may be costly to the supplier.

Production Control has the most data-intensive role. Its effectiveness hinges on the maintenance of accurate inventory data through production transaction processing, supported by visible management and cycle counting, as well as accurate bills of materials and capacity data through engineering.

The **Master Production Schedule (MPS)** does not appear explicitly in Figure 9.4 because, in the organization shown in this picture, it is used within Production Control. The MPS is the list, day-by-day, of the items the company plans to produce, with configurations, quantities, and dates. The MPS does not contain assignments to production lines or the timing of any specific operation within a day or a shift, which is the scheduling task. Production Control generates the MPS and then turns it into several documents that other departments act on:

- The *Materials Requirements Plans* go to Purchasing to be turned into direct orders or into "blanket production orders" that specify a method for triggering actual orders, such as kanban loops.
- The *Timetable/Leveled Sequence* tells production the list of products to start and at what time.
- The *Available to Promise list* tells Sales what they can commit to with customers.

Sales

In Sales, routine processing of customer kanbans and Electronic Data Interchange (EDI) orders can be automated, but the negotiation of terms and conditions and order priorities cannot be. Once an order is on the shop floor, it is no longer just "data," but then involves materials, tooling, and people committed to its execution. As long as it is only data, Sales can move it to the back or the front of the queue and negotiate fulfillment lead times or phased deliveries on exceptional orders as it sees fit for the best interest of the company.

Purchasing

Purchasing executes sourcing of materials needed to meet production plans. The input from Production Planning and Control is the material resource plan, which details type, quantity, and timing of parts needed. In order to achieve better purchasing rates or conditions, Purchasing may enter negotiations with suppliers or order larger quantities to earn discounts.

The pattern in Figure 9.4 is complex, and many factories simply attempt to respond to the daily flow of orders without giving their suppliers any heads up about the future or even anticipating it for themselves. The companies that try to cope with it simplify the problem by doing one of the following:

- Consider aggregate volumes for product families.
- Rely on the management structure to provide accurate data.

The question from Sales may be formulated as, for example, "Can you increase production from 1,000 to 1,200 motors/day?" The plant manager then passes the question on to product line managers, assuming that being responsible for implementing the change will motivate them to provide workable answers. The end-result is a series of what-if analyses performed by product line engineers usually on electronic spreadsheets, and occasionally with production simulators.

> **Stop and Think 9.1**
> Think about how you plan your grocery shopping to "match supply with demand." Do you know anyone who is more systematic about their grocery planning than yourself? What are the pros and cons of grocery planning and how do you think those lists apply to manufacturing?

9.3. Leveled sequencing

Leveled sequencing is used in mixed-flow assembly to smooth the incoming flow of materials and thereby dampen the bullwhip effect. It has no direct impact on customer experience. Of course, customers eventually benefit, indirectly, from improvements to inbound logistics.

Let us assume, for example, that the order fulfillment lead time for a consumer product is 5 days, that final assembly is sequenced every shift, and that products are shipped directly to consumers by currier. A customer will see a consequence if a unit is assembled one day or the next, but not if it is assembled at 9:00 AM or 11:30 AM. Short-interval sequencing does not affect the date and time of delivery into the customer's hands. On the other hand, it does affect the operations of final assembly, internal subassembly lines, and suppliers, particularly those who use body-on-sequence.

9.3.1. Leveled production

Orders from multiple customers for multiple products fluctuate. Sometimes the peaks exceed the capacity of the line, and at other times, there is ample slack. While manufacturing is about responding to customer demand, the approach is *not* to pass the unfiltered flow of orders straight through to the manufacturing floor. Much massaging and filtering of customer demand is needed to schedule final assembly operations in a way that supports pulling parts from the supply chain.

Leveled production is particularly a goal of lean manufacturing and is known as "heijunka". Figure 9.5 shows the intended result. On the right side, firm orders for regularly made items are smoothed into daily rates and supplemented with forecasts for the longer, with time set aside for "Other" demands.

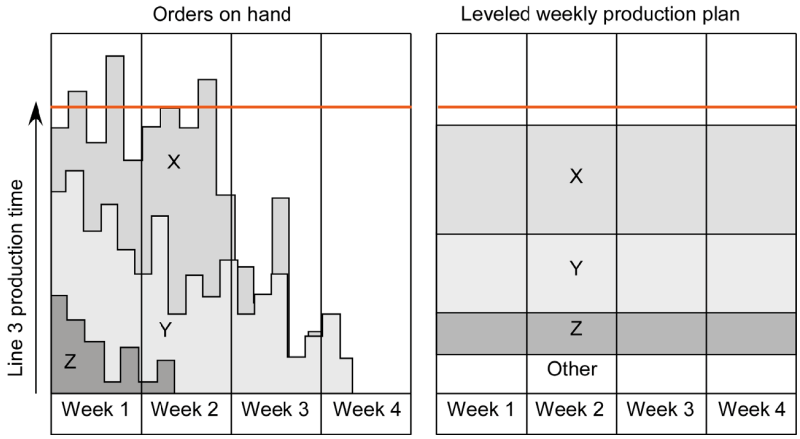


Figure 9.5 From flow of orders to leveled plan

Generating and updating a leveled production plan with 20 production lines and 500 products and not just for the next four weeks but for the next three to six months is a daunting task. But it must be taken on in some fashion in order to support such decisions as:

- Adjusting work time, by adding or cancelling overtime, or changing the shift pattern.
- Acquiring new equipment or setting up new production lines.
- Providing suppliers with input to their own planning process.

9.3.2. Don't shuffle orders already in production

Let us assume for a moment that you are running a tightly balanced line with no slack capacity. The approach to accommodating rush orders from favored customers is to shuffle the orders prior to releasing them to the shop floor. Once released, no shuffling is possible, and they should be thought of like passengers on a chair lift.

The orders may be broken up into small lots or individual pieces to make their way through one-piece flow lines and FIFO buffers to come back together as a customer order when finished. The hand of production control is not allowed to disrupt the flow of materials on the shop floor, which has been engineered to be fast.

This works when the flow time through the shop floor is short. The only reason you have expediters switching jobs around in traditional shops is that the lead times are long. If it takes three months to put three hours of work on a part, then moving it to the head of the line at each station may cut this down. On the other hand, if it takes one shift to do it for every part going through, then there is not much at stake in switching them around on the floor. If the plant is running at full capacity, orders may have to wait, but the only place they are allowed to wait is before release, like skiers in front of the chair lift. At that point, they are normally data only, and can be shuffled to suit the needs of the Sales department, within the constraints of part availability and line capacity.

9.3.3. Don't build incomplete products

A common practice in factories with long production lead times is to start production on products for which some components are currently not on hand, and may not even be expected to be on hand by the

time the product reaches the point in the process where they are needed. Units that are "99% complete" but cannot be shipped wait in a "final check" area until the missing components arrive. Management's expectation is that, as a result, the units will be completed and shipped earlier than if they had not been released for production until all the components were available.

As in the previous section, any possible advantage that might have been gained is negligible once lean manufacturing has reduced production lead times from weeks to hours. On the other hand, the following disadvantages of this approach remain:

1. It increases production lead time and WIP.
2. It commits common parts early, preventing them from being used on units of other products that you could have shipped, sometimes resulting in deadlocks where two incomplete units each contain the parts that would enable the other one to be finished.
3. It complicates logistics, causing the same parts to be delivered for the same use either to production or to the "final check" area.
4. It causes production work to be performed in repair mode, away from the controlled environment of the production line, with its instructions, fixtures, and tools. As a result, not only is the work done inefficiently, but its quality is jeopardized.

With very few exceptions, the practice is not allowed in a lean environment. When faced with the dilemma illustrated in Figure 9.6, the clear choice is to wait until all parts are available.

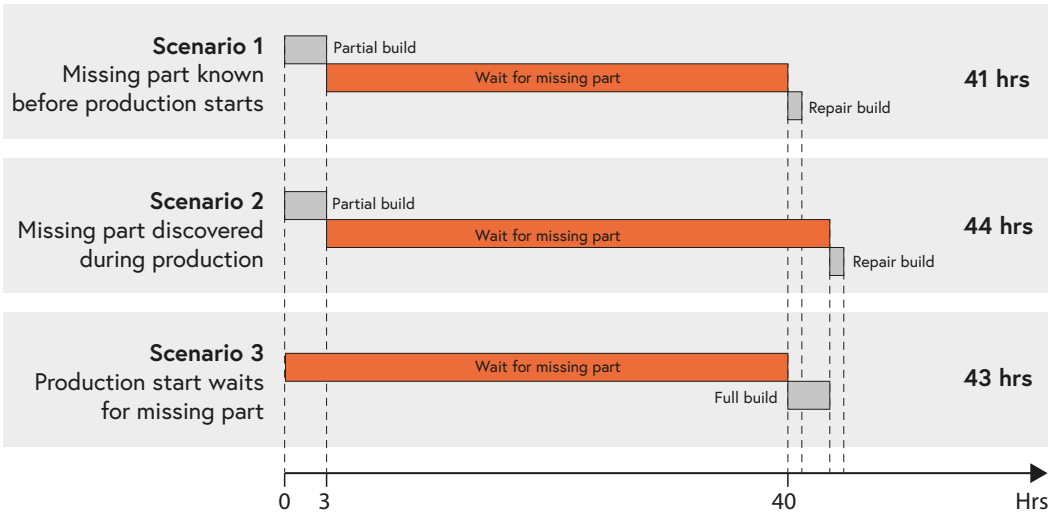


Figure 9.6 Building incomplete products versus waiting for full part availability

Figure 9.6 shows three possible scenarios, in all of which the time needed to complete the product unit is dominated by the 40-hours waiting time for the missing part. The fastest completion is Scenario 1, with the shortage known ahead of production start, where the combination of partial build, waiting and repair build results is shown as finishing the product two hours earlier than the alternative, after a 40-hours wait for the part. It is unlikely to affect customer experience and not worth pursuing, in light of the problems listed above.

Scenario 2, however, is more common in factories that choose to do partial builds with missing parts, as faulty bills of materials and inaccurate inventory data often cause shortages to be discovered only when

assemblers reach for a part and fail to find it. In this case, the wait for the missing part does not start until the partial build is finished, and the product is eventually completed later than in Scenario 3, where the production start is postponed until all parts are available.

The few exceptions to the prohibition of partial builds include car factories, which do not stop production because of delivery glitches on easily accessible, non-safety related parts like ornaments or door panels. On the other hand, production is immediately stopped in response to shortages of parts like proportioning valves, which allocate brake fluid between the front and the rear of the car and are inside the structure.

9.3.4. Reduce setup times to reduce batch sizes

The best way to ship daily to a customer, is to make every day what goes out every day. Long setups, however, make that impossible by forcing production runs to cover multiple days. Also, poor process quality hinders manufacturers from producing only what is needed in a day. Due to long setups, large batches of products must be stored in a finished goods warehouse, from which daily shipments are drawn, as shown in Figure 9.7.

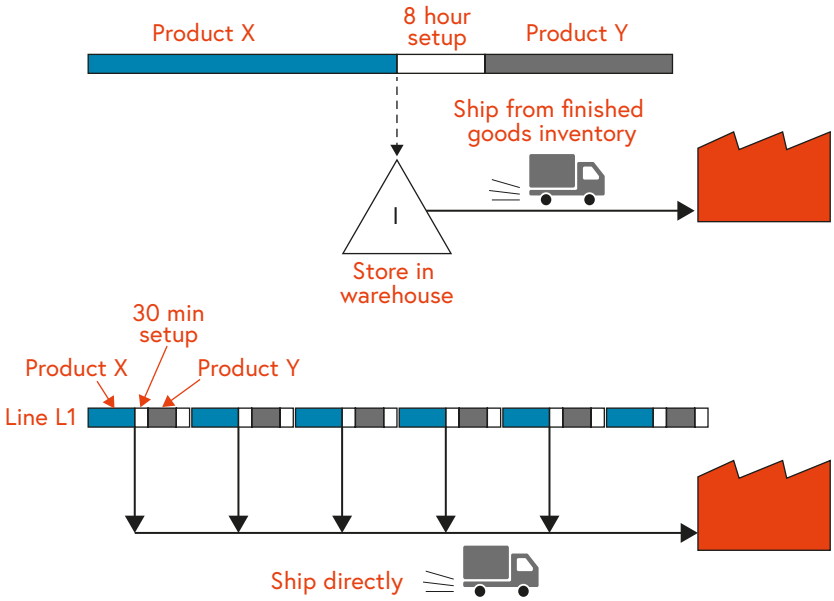


Figure 9.7 Quick setups and leveled sequencing

Classic texts on Operations Research (OR) like Hillier and Lieberman's[3] do not discuss setups at all. Even Salvendy's IE Handbook[4] barely mentions the value of reducing setup times. Detailed coverage of methods to do it is only found in translations of Japanese texts like Arai and Sekin's *Kaizen for Quick Changeover*.[5] One of the key points of the Toyota Productions System is the refusal to accept long setups as inevitable. SMED projects can reduce 8-hour set-ups to 30-minutes or even 6-minutes setups, which

3 Hillier, F. S. & Lieberman, G. J. (2015). *Introduction to operations research*. McGraw-Hill.

4 Salvendry, G. (Ed.). (2007). *Handbook of Industrial Engineering*. John Wiley.

5 Arai, K. & Sekine, K. (2006). *Kaizen for quick changeover: Going beyond SMED*. Taylor & Francis.

enable the more desirable sequencing pattern shown at the bottom of Figure 9.7. But setup time reduction, however diligently pursued, is not instantaneous. While it is going on, and as long as the setups do not vanish altogether, the problem remains of setting batch sizes and sequencing products.

9.3.5. Decide the correct batch size

The Economic Batch Quantity (EBQ) calculations that are traditionally used to set batch sizes are based on assumptions that are not valid for set-ups on production equipment. The EBQ formula is an adaptation of Ford W. Harris's Economic Order Quantity (EOQ) from 1913 to a different context. Placing an order has an identifiable variable cost whose main component is freight. If you have to get a third-party logistics provider to bring a truck of materials from a supplier, then there is a cost for this truck that is spread over its content. If it is half-full, you will spend twice as much per unit as if it is full.

By contrast, for setups, the main resources are machine and operator time, both of which are already paid for, regardless of whether they are used for production or setups. Quarterly or yearly accounting reports can tally the resources spent on setups and deduce an average cost per setup, but this is not a variable cost that is incurred by doing one more setup this morning.

Where setups become expensive is when they are so frequent that they don't leave enough production time to meet the demand and therefore result in lost sales. As shown in Figure 9.8, the marginal setup costs are discontinuous – negligible as long as you leave enough production time to meet the demand, but unaffordable as soon as you don't.

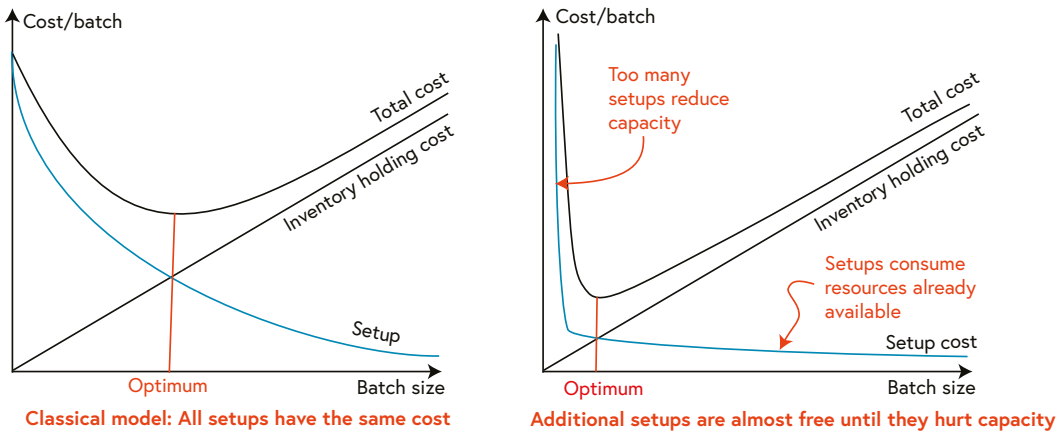


Figure 9.8 Marginal cost of setups

For a machine shop supplying an assembly line, the smaller your batches, the lower your WIP and the better balanced it is in mix. Therefore, the best lot size is the one that allows you to do the maximum number of setups that can be done without eating into the required production time. If all the setups are equal, for any given period, it is a simple calculation:

$$Setup\ budget = Total\ available\ time - Required\ production\ time$$

$$Number\ of\ setups = Setup\ budget / Setup\ time$$

Then you set the batch sizes for your different items to match the needs of assembly. For example, if one unit of the assembled product takes two X's and one Y, then you make the batch size for X twice as large as for Y.

The specific circumstances of each line must of course be considered. For example, if instead of supplying an internal assembly line, a cell makes parts that go directly onto a truck to a customer, in order quantities set by the customer, then there is no advantage in making batches smaller than the quantities going into the truck. In that case, smaller batches would only have to wait.



Figure 9.9 Machine setups can take hours

The constraints imposed on a flexible line by setups can vary:

1. **Equal set-up times.** All setups from any item to any other item take the same time. Then the constraint is on the number of setups that can be performed while still leaving enough time for production. This constrains the size of production runs, but not their sequence.
2. **Large set-up time variations.** Some detergents contain phosphates and others don't. Changing over a spray drying machine is quick when both the from- and the to-product contain phosphates, when neither does, or when the to-product does. On the other hand, changing over from a product containing phosphates to one that doesn't takes 10 times longer because it requires cleaning all pipes. Then you try to perform all the short setups possible before a long one.
3. **Set-up process dependencies.** There is a matrix setup times for every pair of from- and to-product. These setups may even be asymmetric, so that going from product X to product Y does not require the same amount of time as from product Y to product X. This happens when setups involve raising or lowering a temperature or changes in paint color, where light to dark is easier than dark to light. Then there is an exact sequence in which to make a run for every product while minimizing the total setup time. Mathematically, it is the traveling salesman problem (TSP).
4. **Set-up sequence dependencies.** The setup time from one product to the next may depend on the entire sequence of products made since the machine started up empty. This can happen on a machining center when loading tool pockets with the tools missing for the next Product. This is

engineered away by standardizing tooling packages.

Assume, for example, that we have 816 minutes of work time available on two shifts to make products X, Y, and Z, with required production times as shown in Table 9.2. The time available for setups is 816–700 = 116 minutes.

Table 9.2 Production time calculation

PRODUCT	Quantity	Time/piece	Time required
X	100pcs	3.50Min/piece	350Min
Y	150pcs	1.67Min/piece	250Min
Z	75pcs	1.33Min/piece	100Min
Total			700Min

If all setups take 15 minutes, then it is possible to do 7 setups everyday, which takes 105 minutes. The sequence in which the products are made is then not constrained by the setup structure. On the other hand, if we have a product family structure such that changing from X to Y or Z takes 60 minutes while changing from Y to Z or Z to Y only takes 10, then there is a strong incentive to cycle through all the products in one family before changing between families, as shown in Figure 9.10.

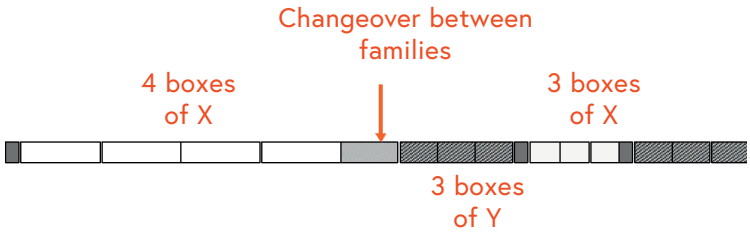
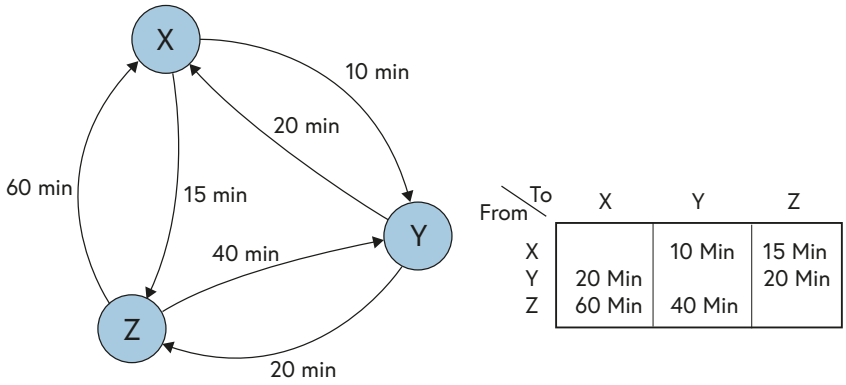


Figure 9.10 Changeover within and between families

The more general case of a full setup matrix is illustrated in Figure 9.11. Then the time spent on setups to go through the entire list of products depends on the sequence. In a more general situation, the setups depend on both the "from" and the "to" product and are asymmetrical: going from X to Y doesn't take the same amount of time from Y to X. Examples include the following:

- If one of the setup parameters is a furnace temperature, then the times needed to raise or lower the temperature are different.
- In painting, going from a light color to a dark color is faster than the reverse.
- In detergents, going from a phosphate-free powder to a powder containing phosphate is easier than the reverse, as discussed above.



From \ To	X	Y	Z
X		10 Min	15 Min
Y	20 Min		20 Min
Z	60 Min	40 Min	

Figure 9.11 Full, asymmetric setup matrix

In the example of Figure 9.11, the impact of sequencing products differently is as follows:

$$X \rightarrow Y \rightarrow Z \rightarrow X \rightarrow Y: 10 + 20 + 60 + 10 = 100 \text{ minutes}$$

$$X \rightarrow Y \rightarrow Z \rightarrow Y \rightarrow X: 10 + 20 + 40 + 20 = 90 \text{ minutes}$$

Sequencing setups in this case is akin to solving the traveling salesman problem. Each product is like a city that the machine needs to visit, the setups are one-way roads with different travel times, and we are trying to minimize the total travel time to visit every city. Operations Research software can find an optimal sequence for thousands of products but real situations rarely involve more than ten or even five products, and can easily be solved with simpler methods.

9.3.6. The every-product-every strategy

The "every-product-every strategy," or "fixed repeating sequence," is particularly useful for rate work, when setup times vary with which products you are changing from and to. This involves two steps:

1. As discussed above, find the sequence with which the line can cycle through all the products with the minimum total setup time.
2. Choose the shortest production run length for the products that are compatible with meeting the demand.

Establishing a fixed repeated sequence is work, but it does not have to be done every day. It pays off in good utilization of the lines doing rate work, and in a mode of operation that is simple and easy for operators to follow. In later sections, this chapter shows how this approach is combined with the pull system to make it responsive to changes and fluctuations.

Figure 9.12 shows a weekly repeating schedule for a line making the same three products every day.

	Mo	Tu	We	Th	Fr
7:00 AM					
8:00 AM			X		
9:00 AM					
9:50 AM					
10:00 AM					
11:00 AM			Y		
12:00 PM					
12:45 PM					
2:00 PM			Z		
3:00 PM					
4:00 PM			Other		

Figure 9.12 Leveled daily production plan

9.3.7. **Leveled scheduling in mixed-flow assembly**

In mixed-flow assembly, setups between products are instantaneous, and successive units coming down the line can be for different products. Toyota does it for final assembly of cars (Figure 9.13), by using sequencing algorithms that smooth the flows of thousands of incoming items. The simplest approach is to spread the shift quantity for each product over the length of the shift and drop these patterns onto a common timeline.



Figure 9.13 Mixed car assembly line in a Lexus factory in Kentucky (Credit: Toyota)

Box 9.2 **Using kanban boards to implement an every-product-every strategy**

A fixed, repeating schedule can be made responsive to fluctuations in the work by treating it like a set of reservations to be confirmed. In timetable planning, they are confirmed by changing the background color on a spreadsheet cell. The communication of this information to the shop floor, however, may require printing and distribution of hardcopy. This can be avoided by using kanbans for confirmation.

Figure 9.14 shows the flow of confirmation kanbans. Each full pallet carries a card. When the pallet is loaded onto a truck for shipment, the card is pulled out and placed in a collection box. These cards then return to the shop floor and are posted in a card pocket on the scheduling board. With this approach, a simple planning logic can be used without jeopardizing the plant's ability to respond to demand fluctuations.

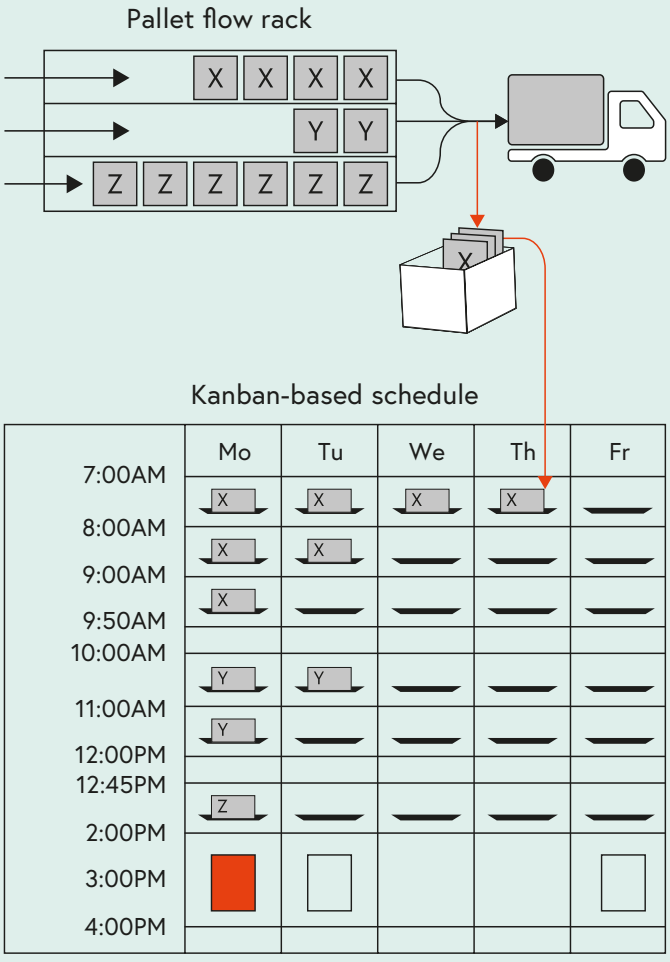


Figure 9.14 Schedule confirmation by kanbans

Strictly speaking, however, this only smooths the incoming flows if the products have no common parts, and products assembled in a mixed-flow line usually have common parts. Toyota's algorithm builds the sequence incrementally, by selecting for the next unit the product that generates the least deviation from a smooth consumption curve for the parts.

Although leveled sequencing is described in the lean manufacturing literature, it is not widely practiced outside automotive industries. Assembly managers often see it as a major change whose value they fail to see, because it targets the upstream supply chain rather than their area of responsibility. There are production lines, most commonly in final assembly, that can change between products with no down-time, and therefore can have production runs of a single unit. Outside of the automobile industry, manufacturers are taking advantage of the opportunities this creates.

How leveled sequencing works with the pull system

The most elaborate form of leveled sequencing takes place in mixed-flow assembly when setups are instantaneous, either because the parts for all products are available on the line side at all times or because they have been *kitted* – picked and collected for a specific product – and sequenced off line.

Figure 9.15 shows the high-level materials and information flows associated with this approach. Shown in black, the black dots represent single pieces moving to the next operation like clockwork at takt intervals and the other symbols represent the storage, transportation, and testing steps the organization has not found a way to eliminate yet.

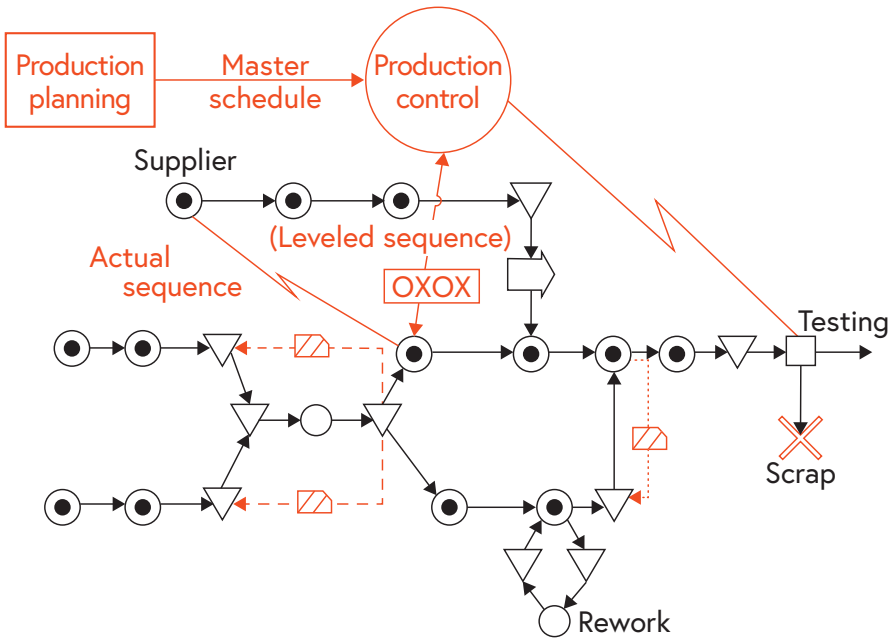


Figure 9.15 Leveled sequencing in mixed-flow assembly

Production control receives daily demand data in the form of a master schedule from Production Planning, and real time updates on actual production from Final Test. It uses this data to generate a leveled sequence with algorithms described below. The leveled sequence is passed to the *pacemaker* station, usually the first station of final assembly, which commits itself to each unit in the sequence by actually starting it.

Communication between Production Control and the pacemaker goes both ways. Occasionally, the first station of final assembly does not have the materials needed to start the next unit in the sequence, in which case, consistent with the principle of not building cripples, it skips to the following one. The actual start sequence therefore has small variations from the calculated sequence. This is why body-on-sequence suppliers produce based on the actual start sequence. Upstream from the pacemaker, the flow of most components is managed through kanbans, and, as we have seen, this system adjusts automatically for small variations in demand.

Stop and Think 9.2
Why are instantaneous set-ups beneficial for mixed-flow assembly lines?

Leveled sequencing method #1

The concept behind the simplest method for leveled sequencing is shown in Figure 9.16. The demand for a product, over one shift, is spread evenly over the shift and then dropped onto the mixed-flow line's common timeline for the shift. It is focused on getting all the products out on their own takt times, and works perfectly if the products have no common parts.

If there are common parts, then, from the point of view of such a part, a sequence of units for products it goes into is equivalent to a run of identical units. Since the primary purpose of leveled sequencing is to smooth the flows of materials upstream from final assembly, a more sophisticated method is needed in the presence of common parts.

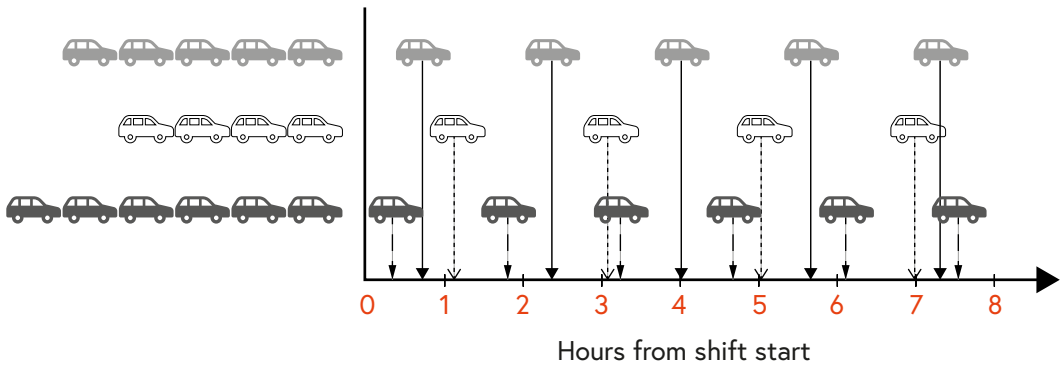


Figure 9.16 Leveled sequencing method 31 concept

Drawings similar to Figure 9.16 are commonly found in the literature, but it is somewhat misleading because cars assembled on the same line do have many common parts, and this method alone would not yield a smooth consumption of these parts. Method 2, below, addresses these issues.

Leveled sequencing method #2

The second method is designed to smooth the consumption of a set of parts. It is iterative as shown in Figure 9.17. For each part, the diagonal line is the ideal cumulative consumption through the shift. The staircase line is the cumulative consumption of each part based on the sequence already decided and on the impact of each candidate to be the next item.

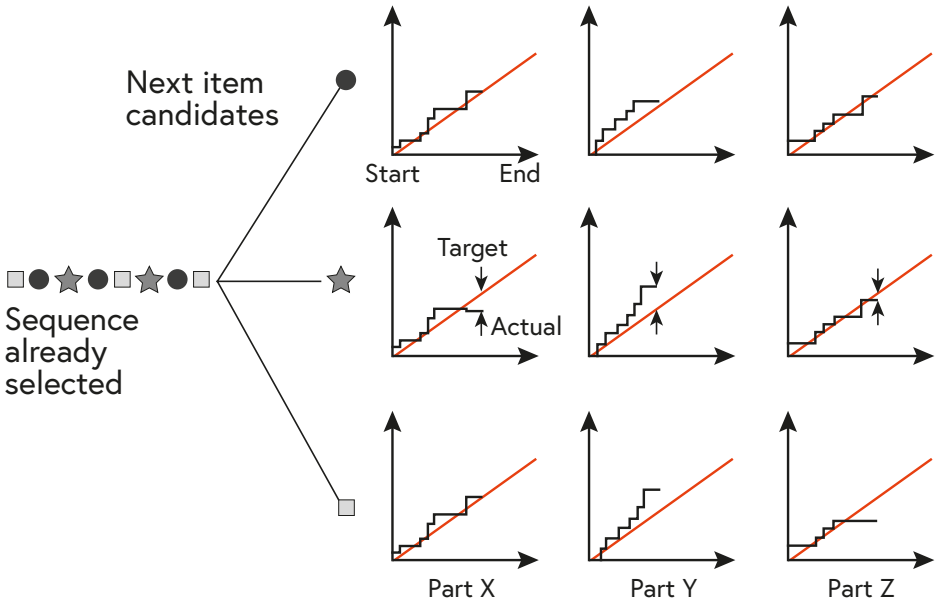


Figure 9.17 Leveled sequencing method #2 concept

For each candidate product to be the next item on the list, examine how far it takes the actual cumulative consumption from the straight line ideal for each part in the set. Then select the product giving the least overall deviation. This algorithm is described in detail in Yasuhiro Monden's *The Toyota Production System*.[6] The measure of deviation from the smooth target line to the actual consumption staircase is the square of the difference, summed over all the parts in consideration, and the selected product can be viewed as a least-squares fit. Over time, Toyota has refined this method by adding rules to control the interval between identical vehicles.

6 Monden, Y. (2010). *Toyota Production System: An integrated approach to Just-in-time*. Productivity Press.

Box 9.3 **Example of balanced scheduling using method 1**

Suppose we are making three products called "Dot," "Square," and "Star," and that the master production schedule is as shown in Table 9.3.

Table 9.3 Master production schedule for three imaginary products

Item	Quantity
Dot	10
Square	5
Star	7

Figure 9.18 shows how to apply method 1 with an electronic spreadsheet. You start with the shift demand by product. Then you list all the units you are going to make, assigning them a sequence number by product. When you divide that sequence number by the demand, it gives you the proportion of the cumulative total demand for each product represented by each unit. We add one to the demand rather than just the demand to keep the last units for all products from scoring 100% and needing a tie breaker. Tie breakers may still be needed if different products have exactly the same demand. Then we sort all rows by increasing values of this ratio, which gives us the leveled sequence. This simple method is used in some automotive and aerospace assembly applications.

Unit	Rank	Rank/(Demand+1)
Dor	1	9%
Dot	2	18%
Dot	3	27%
Dor	4	36%
Dot	5	45%
Dot	6	55%
Dor	7	64%
Dot	8	73%
Dot	9	82%
Dot	10	91%
Square	1	17%
Square	2	33%
Square	3	50%
Square	4	67%
Square	5	83%
Star	1	13%
Star	2	25%
Star	3	38%
Star	4	50%
Star	5	63%
Star	6	75%
Star	7	88%

Unit	Rank/(Demand+1)
Dot	9%
Star	13%
Square	17%
Dot	18%
Star	25%
Dot	27%
Square	33%
Dot	36%
Star	38%
Dot	45%
Square	50%
Star	50%
Dot	55%
Star	63%
Dot	64%
Square	67%
Dot	73%
Star	75%
Dot	82%
Square	83%
Star	88%
Dot	91%

Figure 9.18 Sequencing three products with Excel

9.4. Autonomous shop floor pull systems

The *pull system* eliminates the need to direct the work of the entire plant from a central location, as is attempted with the *push system* of MRP in ERP. Centrally scheduling hundreds of processes and machines with different setup structures, load sizes, and modes of operation is practically very difficult. Hence, if complexity can be reduced by using simpler rule-based pull systems in parts of the production system, manufacturers have incentives to do so. By regulating the high-level flows between lines, departments and plants, the pull system reduces the challenge to that of locally scheduling tens of units of a few machines or stations each. As long as each of these units delivers the demand expressed by the flow of pull signals, they simply follow their own local scheduling rules.

Pull methods can be better tailored to local needs than if a single push method had to apply to the entire plant. In the simplest case of a cell dedicated to one product, picking kanbans first-in-first-out from a box may be all that is needed. At the opposite end, an Advanced Planning and Scheduling (APS) system may be justified to schedule work through the kind of Flexible Manufacturing System (FMS) that are used to manufacture components used in aerospace or semiconductor production equipment applications.

Marked locations, lights and drop-down plates can be used as pull signals provided the respondent has the required materials on hand when the device is visible. By contrast, fixtures, pallets, or bins can travel, and, therefore, when used as a pull signal, have a range that is not limited by the line of sight from the destination point, but their bulk and weight limit information handling.

This restriction is lifted by using *kanbans* – that is, cards, tags, or other small and mobile physical tokens used as pull signals and carrying only information. Kanbans can be sequenced and posted on boards and can carry bar codes as well as human-readable data.

9.4.1. Location labeling

Empty marked locations, lights, or drop-down plates are all devices that can be used to signal a need. Such signals, which are easily understood by all human workers, can be very quick and cheap to implement.

Their common limitation is range: They cannot be perceived and acted upon beyond the line of sight of the devices. In addition, the agent who sees the signal must be equipped to act on it. If, for example, a materials handler who notices that a tote of item X is needed at cell A does not have such a tote at hand, he or she cannot act on the signal. Writing down the item requirement in a shopping list for bringing a tote from the warehouse is not acting on the signal but creating a new one. As a consequence, the use of this type of signal is restricted to items consumed at a sufficiently steady rate to make it feasible for the handler to always carry a spare tote or bin on milk runs.

With cells that are in close proximity, an empty space on the floor with proper markings can be an effective pull signal. It helps implement "take one, make one." This is the way actual supermarkets work: Employees keep shelves full, and replenish any empty space they see. For this to work as an alternative to kanbans, operators at the upstream operation must be able to see the input buffer of the downstream operation. Between operations that both practice one-piece flow, this space can be just a painted shape on the floor. It does, however, need more structure when the upstream and downstream operations have different batch structures.

Figure 9.19 shows one example from a construction equipment factory. Note that the white painted circles match the part and the mobile stop signs are used to limit WIP.



Figure 9.19 Visual location labels are intuitive pull signals; here used in a construction equipment factory

9.4.2. **Two-bin systems**

The two-bin system is the grandfather of all pull systems, but still appropriate in the right circumstances. Two containers are used for the item, and an empty container is a replenishment request, as shown in Figure 9.20. The system works as long as a full container is delivered before the other one is empty. They must therefore be dimensioned so that one bin covers the production needs during the replenishment lead time. But if it doesn't, the system can be extended with a third bin.

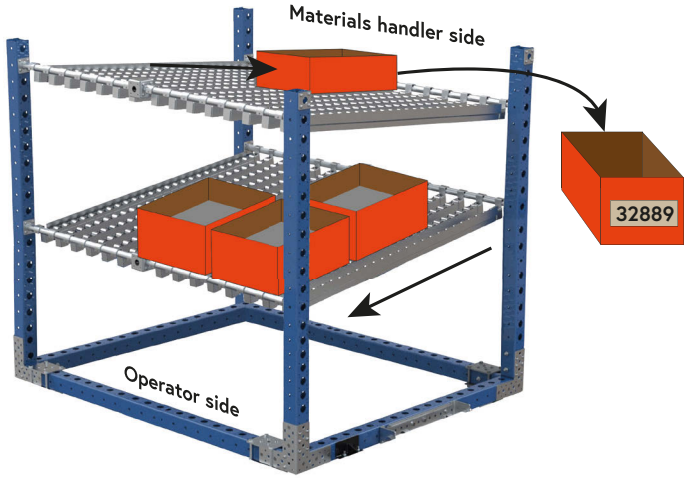


Figure 9.20 The two-bin system implemented with flow racks

The containers must be marked with all the item and shelf location information required for proper routing. The container is dedicated to the item, but this is true anyway for all the items for which there are slots and dunnage in reusable containers.

The two-bin system constrains the handling of empty containers. If the containers serve as pull signals, then any empty bin must directly return to where it can be filled. If empty containers are just empty containers, they can wait, be stacked and picked up by a separate service at a much lower frequency than pull signals.

Also, with this method, the pull signal cannot be issued until the container is empty, whereas a kanban is pulled as soon as the first part is taken from the container.

9.4.3. **Empty fixtures or pallets**

The flow of materials between operations can sometimes be regulated by controlling the number of a particular type of device in circulation. Figure 9.21 shows a machining process that involves roughing and finishing on two different machines that are not collocated. The obvious solution of bringing the machines together into a cell is not feasible because the present situation is not expected to endure long enough to pay back the cost of moving the equipment. A new machine is scheduled to arrive within six months that will do all cuts from raw casting to finished part.

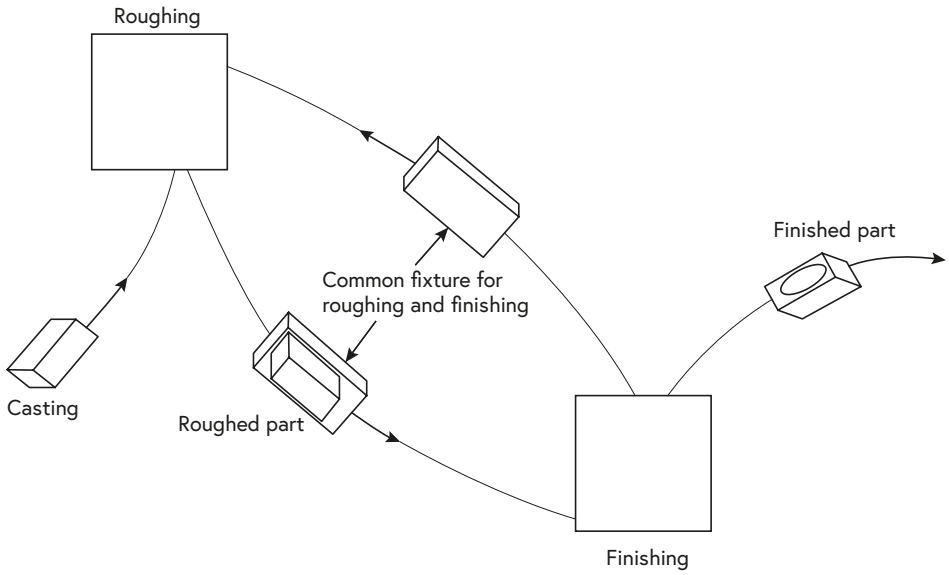


Figure 9.21 Using fixtures as pull signals

If both the roughing and the finishing operations can be engineered to use the same fixture, it not only saves the cost of taking parts on and off between operations, but it turns an empty fixture into a pull signal that cannot be ignored. It is physically impossible for the roughing operators to start a part unless they have a fixture to hold it.

9.4.4. **Kanban**

The approaches we have discussed so far either provide signals with a range limited to the line of sight from the destination or with severely limited information-handling capabilities. The key to freeing ourselves from these restrictions, where needed, is to separate the information from devices used to handle the materials.

Information today is thought of primarily in electronic form and traveling across computer networks. 50 years ago, when Taiichi Ohno and Kiichiro Toyoda were developing the TPS, it meant hardcopy carried by vehicles and people. Today, the need for hardcopy has vanished for some functions, but it is easy to forget that is has not for others. In particular, as long as people are involved in production or logistics, they will benefit from human-readable documents attached to parts or containers.

While Toyota no longer uses the pure hardcopy system, it still applies the same logic. "Kanban" is a common everyday Japanese word meaning: Sign, doorplate, or shingle, as shown in Figure 9.22. As we shall see, in production control, the meaning is quite different. Everyday life kanbans don't usually move; shop floor and supplier kanbans do.

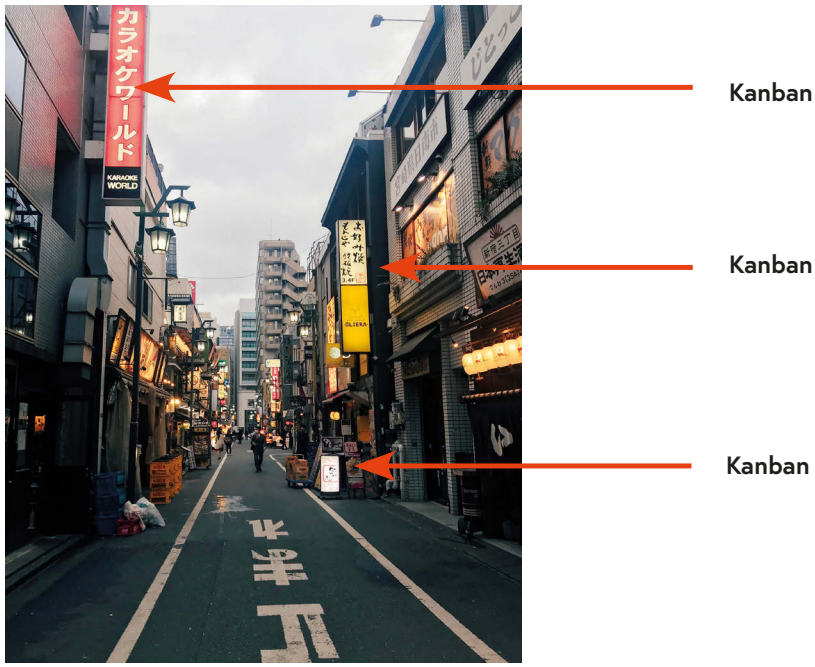


Figure 9.22 Kanbans in everyday Japanese life means signpost

In manufacturing, a kanban is a physical token used as a pull signal. What separates a kanban from the signals discussed above is that it is small, mobile, and carries only information. They exist in many variations, and are used both in-house and with suppliers. "Move kanbans" authorize transfers of parts; "production kanbans" authorize the processing of parts; "common kanbans" authorize both processing and transfer; "tunnel kanbans" authorize processing through a sequence of operations.

The kanban system is based on the assumption that the next few hours will be like the last few, which is true in the automotive industry but not of manufacturing in general. While more sophisticated forecasting schemes are not possible with hardcopy, they are with computer systems.

As in the example of Figure 9.23, kanbans are most commonly cards of varying sizes, but, on occasion, metal plates, golf balls, or other creative signal carriers have also been used. While many call any kind of pull signal, or even a shelf, a "kanban," the discussion of these matters is better served if we reserve the term for its original meaning of a physical token, usually a card, used as a pull signal. A kanban and an empty bin are both pull signals, but the card is lighter and smaller than the bin, can be moved, replaced, and changed more easily, and can be mounted not only on containers but as well on boards as part of visual management.



Figure 9.23 Supplier kanban from the Toyota website (Source: Toyota)

A kanban carries all the information needed to act. As in Figure 9.23, it is often in both human and machine readable form. It specifies what should be made, moved, or shipped, from and to which locations, when, and in what quantities. This information not only tells operators, materials handlers, or shipping and receiving clerks what to do but also provides guidance in the execution of these actions, replacing previously existing paperwork, such as dispatch lists, travelers, routing slips, or order slips.

That a kanban carry information that is indispensable in daily operations is key to the success of the system. Many plants initially put kanbans in circulation in addition to the existing paperwork rather than as a replacement for it. This is a mistake for two reasons:

1. It creates more work and adds complexity. To be successful, the kanban system must make the work simpler, not harder, for all those who have to handle the cards.
2. It allows operations to continue with missing kanbans. The kanbans are not indispensable, and can therefore be ignored without penalty.

In the plant, kanbans move according to rules, as a result of operational activity, and these moves, in turn, cause more work to be sequenced. As a result, a kanban loop is self-sustaining, and eliminates expediting. Making the kanban system work requires effort on the part of the supervisors, but at the level of making sure that each loop makes senses and that operators follow the rules. He or she does not have to look for parts.

In addition to its immediate operational role, the kanban system is also a tool for driving process improvement. By quietly removing one or two cards from a given loop, supervisors test the practical limits of WIP reduction. When they "succeed" in causing shortages, it points out areas of potential improvement.

Common kanbans

A *common kanban* authorizes both production and movement. It is discussed first, because it is often the first choice of initial implementers. It is an appropriate choice in many cases, but not always, and the implementers need to know that there are alternatives.

Figure 9.24 shows its application in a plant where a fabrication shop makes components that are each assembled with purchased components into many different finished goods. The fabrication process may be injection molding, plastics extrusion, diecasting, or a variety of other processes that are performing on a set of parallel and interchangeable machines or lines.

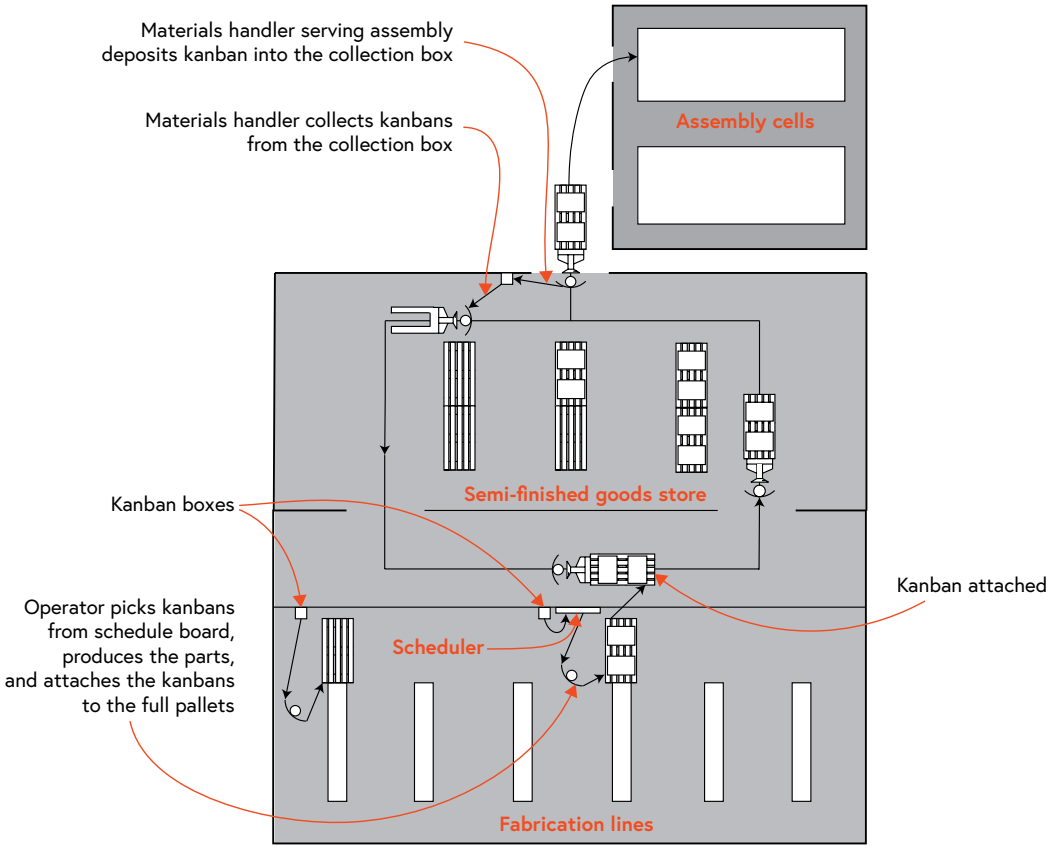


Figure 9.24 Use of a common kanban

The fabrication shop's policy is to make the components to stock and assemble them to order. It has decided to use just one type of kanban and have assembly's pull of a pallet from the semi-finished goods store to trigger both the fabrication of a replenishment pallet for the same component and its transportation to the semi-finished goods store.

The way retrieval from the semi-finished goods store is organized is not our topic here, and we just assume that it is in place. The only feature of it that is relevant to our discussion is the obligation placed on the materials handler, who serves assembly, is to detach the kanban attached to a pallet on its way to the assembly cell and deposit it in the collection box.

The materials handler in charge of replenishing the store "picks up the mail" from the collection box and delivers it to fabrication. The common kanban may look as the example in Figure 9.24. It shows both the production line that is to make the part and the store address to which it is to be delivered. The "delivery cycle" number tells the production operator when the materials handler will need to pick up the full pallet.

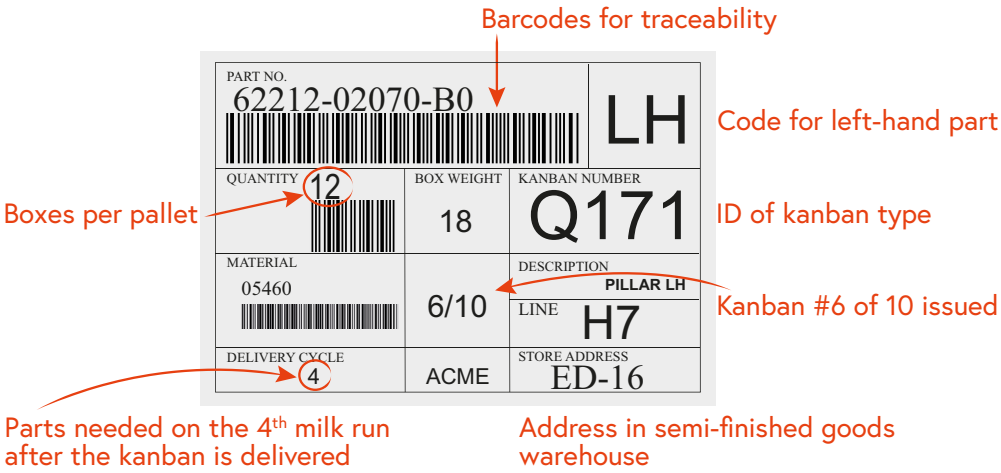


Figure 9.25 Example of a common kanban

Some but not all kanbans are serialized as the one in Figure 9.25. It is marked as number 6 of 10 issued. From a strict production control point of view, this number should not matter: As long as we have one card authorizing an action on a given quantity of parts, it shouldn't matter which one. If the cards, however, are acted on in order of their serial numbers, then, once attached to bins, they represent the order in which the bins were made. The full bins arrive at the destination point with kanbans numbers 1 though 10, and then back to 1. As long as not all the bins are at the destination point, as shown in Figure 9.26, the kanban serial numbers provide unambiguous information about the process sequence, which is useful for traceability and quality control.

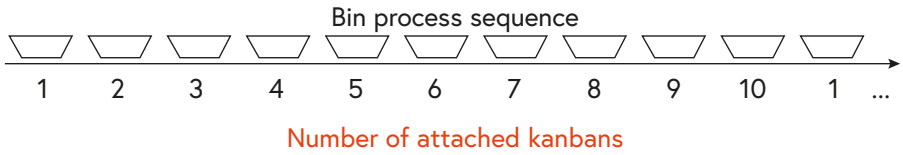


Figure 9.26 Bins with serialized kanbans

For fabrication lines that are dedicated to one item, the kanbans may simply be dropped into a box, to be acted upon in sequence by the production operator. Figure 9.27 shows an example of a simple kanban board in a boat engine factory in Lexington, Tennessee. The board is only for the exhaust pipes supplied to the main assembly line. The line is fully occupied making just this item – or it wouldn't be dedicated to it – and has no opportunity to produce a pallet earlier than needed. The role of the kanban flow to this line is to slow it down or stop it if consumption drops. As can be seen from the orange kanban cards in Figure 9.27, this assembly line is scheduled to make the three exhaust pipes.

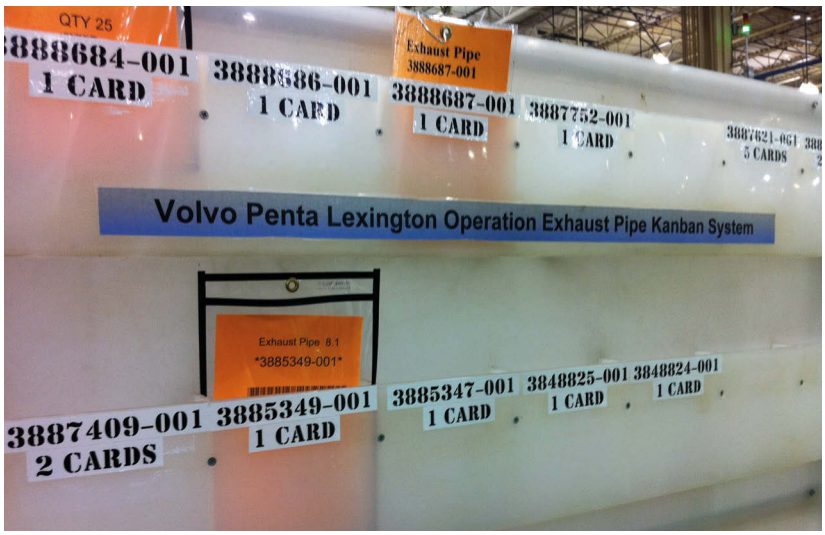


Figure 9.27 Example of kanban board in a Volvo Penta factory in Lexington, TN, USA

The processing of the kanbans is more complex at lines that make more than one item, because fabrication lines have setups that constrain the size of production runs. Setup time reduction eliminates this problem, but not instantly. Projects in this area may have reduced setups from 2 hours to 45 minutes on all the machines, but it will take another 2 years to get them under 10 minutes, and another 5 years to make them instantaneous.

In the meantime, it is common for a line that is assigned, say, 5 different items, to have insufficient capacity to make one-pallet runs in the order the kanbans are received. A scheduling board is then used to organize the flow of incoming kanbans into a sequence of multi-pallet production runs. If it is simple enough, the materials handler can place kanbans directly on the board. The complexity of the scheduling board, however, often requires the intervention of a specialized scheduler. Then it is time to consider some of the different approaches discussed below.

Box 9.4 specifies kanban rules used at Toyota.

Box 9.4 **Toyota's kanban rules**

How kanbans are used is what makes these pieces of paper into instruments of production control. The rules Toyota developed for the use of kanbans should be studied and not tampered with lightly. They are the protocol operators and materials handlers must follow in routine operations. They specify the following:

- The kanban is pulled from a bin when the *first* unit is withdrawn
- Replenishment is for the quantity marked on the kanban
- Kanbans for the same items are processed first-in-first-out
- Kanbans for multiple items may be sequenced on boards
- Kanbans should never be knowingly attached to defectives
- Limiting the number of kanbans makes the process more sensitive and reveals improvement opportunities

The dual-card system with production and move kanbans

For the common kanban, the same kanban is moved by a materials handler, sequenced by a scheduler and produced by an operator. It is less likely to be damaged or misplaced if it triggered only one action by one person, instead of three actions by three agents with different bosses. This is one reason for separating *production kanbans*, which authorize only making the parts, from *move kanbans*, which authorize only their movement between two locations. This separation describes a *dual-card system.*

The dual-card system with different kanbans to authorize production and transportation is less popular because it appears to be more complicated, requiring two types of cards instead of one. The complexity that it adds, however, is in management. In execution on the shop floor, it actually makes the system simpler when dealing with flexible lines and production runs that are multiples of the quantities moved and that vary by product.

The differences with the common kanban system are shown in Figure 9.28. What happens between the semi-finished goods store and the assembly cells is unchanged from Figure 9.24, and therefore omitted. What happens at the end of the fabrication line, on the other hand, is quite different, as shown in the enlargement.

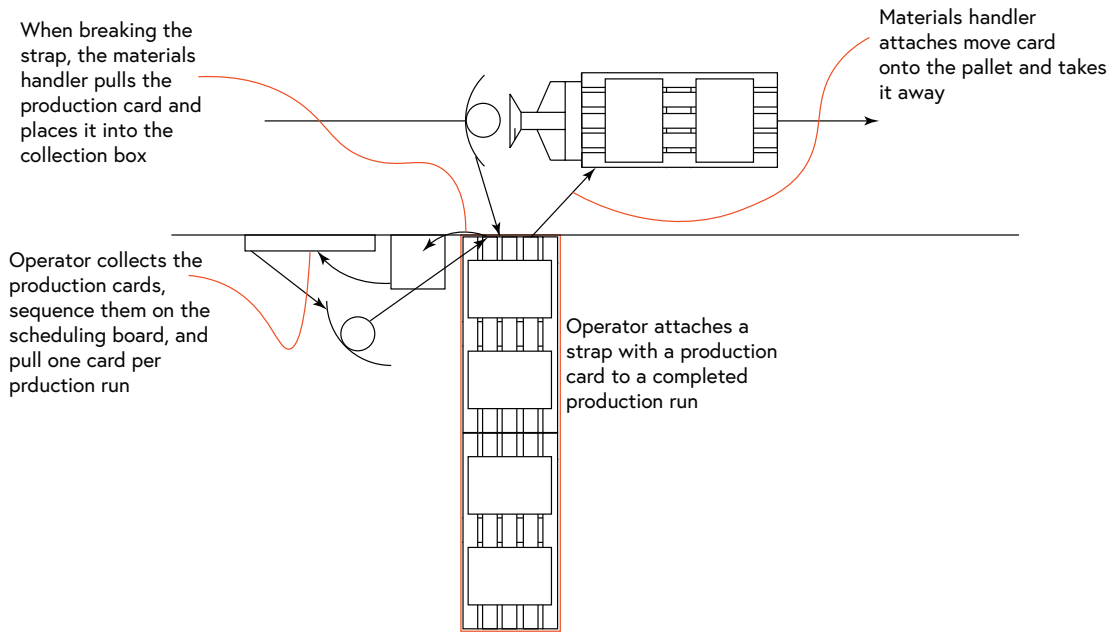


Figure 9.28 Using the dual-card kanban system

The production operator only touches the production kanbans, and there is exactly one for each production run, and, for each item, the number of pallets in the production run is written on the kanban. When completing a production run, the operator puts a strap around the pallets comprising the run and hangs the production card on this strap.

The materials handler may pull a pallet from a partial production run by attaching a move kanban, also known as "transportation kanban," to it and taking it away. To break a new production run, the materials handler has to remove the strap and dispose of the attached production kanban, by placing it into

the collection box. Then, he or she attaches a move kanban to the first pallet in the run and takes it away. The scheduling board is designed so that the next production run is finished shortly before the last pallet of the previous one is taken away.

At the cost of having two types of kanbans instead of one, this approach addresses the issues raised in the previous section:

- **Responsibility.** The materials handler is responsible for the transportation kanbans; the production operator, for the production kanban. Each kanban has a simple loop, and the two loops intersect where the handler detaches the production kanban to place it into a collection box.
- **Posting on a scheduling board.** Since there is exactly one production kanban for one production run, the issues of mapping a stack of kanbans to a production run go away. As a result, a simpler scheduling board can accommodate more items, and the need for a specialized scheduler to organize the kanbans on the board goes away.
- **Consistency.** The handler only takes away a pallet with a corresponding transportation kanban in hand. The operation is therefore completely consistent with a pull system.

Another advantage of using both production and move kanbans is that it decouples improvement efforts on both sides, since the number of move kanbans in circulation or the size of the production runs can be reduced independently. While the common kanban only allows you to control the total amount of material in fabrication and the semi-finished goods warehouse, it does not enable you to control each one separately.

Figure 9.29 shows how the information on the common kanban is split among the move kanban (left) and production kanban (right). Since each production kanban is for multiple pallets in this example, there are fewer in circulation than move kanbans.

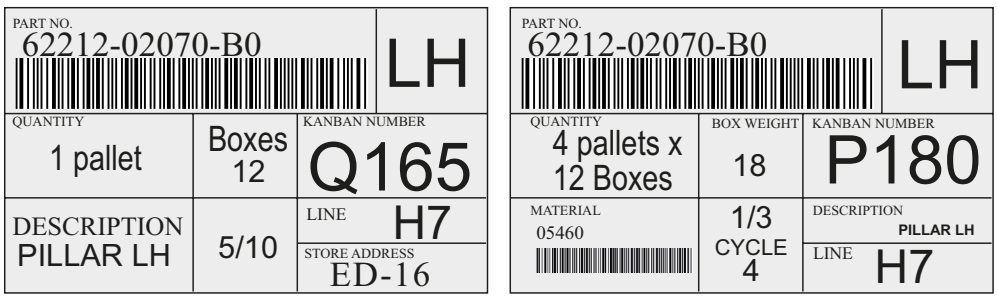


Figure 9.29 Example of the dual system with a move kanban and a production kanban

Stop and Think 9.3

Kanbans have traditionally been physical cards and tokens in factories. What are the disadvantages and advantages of digitizing kanbans into so-called "eKanbans" and using modern IT systems rather than paper-based systems?

Supplier kanbans

Kanbans are not only used to control production within the factory, but also between a factory and its suppliers. Today, while the logic of the supplier kanban system is used in lean supply chains, the stack of recirculating hardcopy kanbans delivered to a suppliers by milk runs trucks is a thing of the past.

The differences between an in-house move kanban and a supplier kanban are due to the need for it to be used by two different companies, between which it serves as a commercially binding order. Incidentally, in 1950, the Japanese tax authorities required documentation on commercial transactions that the circulation of kanbans did not provide. Today, *electronic* supplier kanbans are triggers within the framework of a blanket purchase order specifying terms and conditions.

For illustration, we show in Figure 9.30 an example of an elaborate hardcopy kanban that circulated between Shimizu and Hino in the automobile industry in Japan. Such kanbans, in the form of *one-time use* hardcopy, travel with parts, and the shipping and receiving transactions are recorded by scanning barcodes.



Figure 9.30 Supplier kanban between Shimizu and Hino

The customer's and the supplier's item numbers may or may not match. If they don't, then both item numbers are needed on the card. The item nickname, also called "item back number," is the short name that operators use to refer to the part among themselves instead of the tongue-twisting, confusing full item number. The "1-2-2" on the bottom left describe the delivery cycle as follows:

- The "1" means that the parts are delivered within *one* day of the kanban being issued.
- The first "2" means that a truck delivers parts from this supplier to the customer plant *twice* a day.
- The second "2" means that a kanban leaving on one truck comes back with parts not on the next truck but on the one after it.

Tunnel kanban

There are circumstances where it is practical to control the total amount of materials within a segment of a process without controlling the flow between every pair of operations or lines. Imagine, for example, that you machine parts in-house, send them to an outside contractor for electroplating, and bring them back in-house for painting. Then you may want to use kanbans between Painting and Machining without having any action taken on the kanbans while inside the electroplating "tunnel." Kanbans travel and stay with machined parts throughout their journey to the contractor and back, to be pulled at Painting and returned to Machining. The tunnel kanbans cap the total number of parts in the plating loop as a whole, but do not affect where in that loop they may be. This concept is illustrated in Figure 9.31.

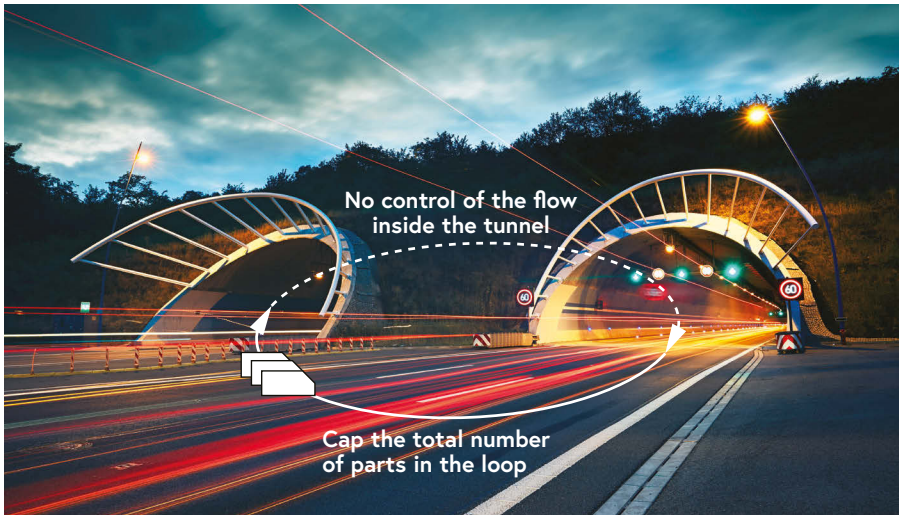


Figure 9.31 The tunnel kanban concept

This concept of capping WIP without controlling where it is on the floor can also be used to manage the residual job shops that remain after a plant has been converted to lean manufacturing. These job-shops typically perform the following function:

- **Production of C-items** – that is, the large number of products that are made sporadically and in small volumes, accounting in aggregate for possibly 5% of the production volume but 80% of the product catalog.
- **Production of fixtures or refurbishment of tools or dies.** A standard process exists only at a high level, and the detailed routing of parts through machines vary with each unit.
- **R&D and pilot production.** Since the work performed is all experimentation, by definition the work cannot be standardized and is non-repetitive.

In these circumstances, restricting the aggregate amount of WIP in the shop avoids congestions and lets jobs proceed quickly through whatever ad-hoc route they may follow. Inside the shop, the work can be sequenced visually as in the example of Figure 9.32, for a machine shop that makes custom jobs with parts small enough to fit, together with their paperwork, tools, fixtures, and NC programs on a cafeteria tray.

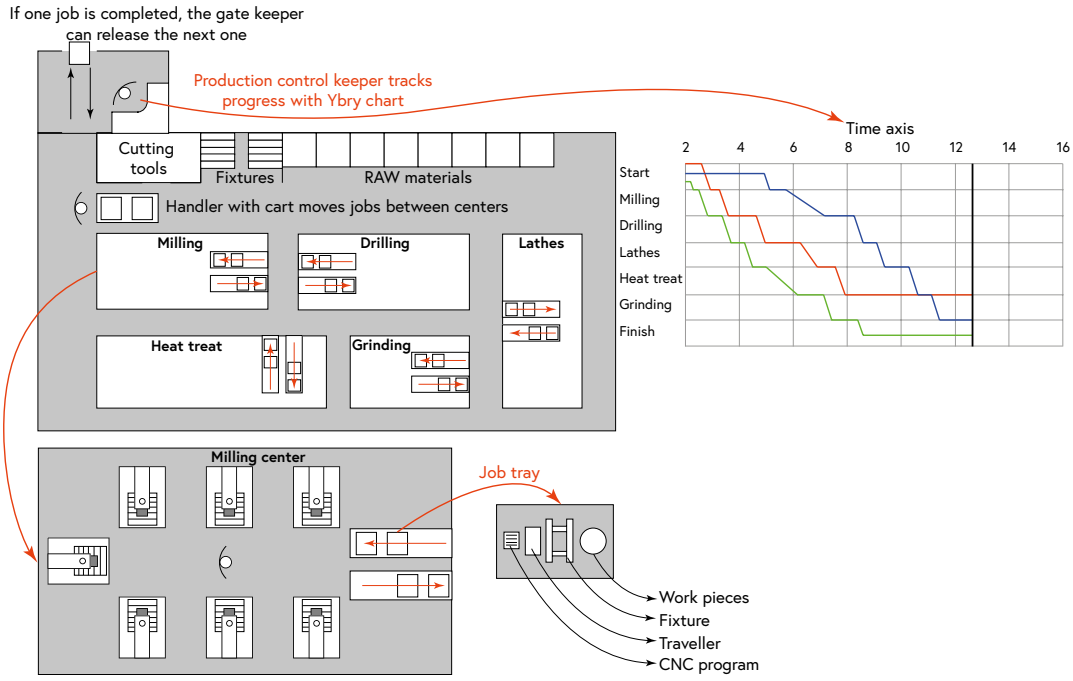


Figure 9.32 Visual FIFO sequencing inside a machining job-shop

The policy is that, say, no more than 50 jobs are allowed in the shop at any one time. If 50 jobs are in process, then the next one waits until one job completes. The high-level common process is that, for example, milling is done before heat treat, and grinding afterwards, and this high-level process is used to map the progress of all 50 jobs in process on an Ybry chart.

All the job shuffling and priority games are played by the gatekeeper before a job is released. Once released, it proceeds first-in-first-out (FIFO) in every center specified on its traveller through gravity flow racks. Since the WIP is capped at a low level, all jobs complete quickly once they are released. In Figure Y, the milling center is laid out as a pseudo cell. Since it is comprised of equivalent machines working in parallel, it is not a cell, but the layout borrows as much from cells as possible.

> **Stop and Think 9.4**
> During the COVID-19 pandemic, many retail stores implemented a system where customers were required to use a shopping cart when entering the store. The stores capped the number of carts. What kind of kanban system is this? What is the purpose and benefit?

9.4.5. Other card-based pull methods

There are numerous attempts at creating other low-key card-based shop floor control systems. The most common alternatives to kanban in job-shop systems are visual reordering points, CONWIP, and POLCA. Two elaborate books by Matthias Thurer and his coauthors and Christoph Roser review and explain these and a series of other card-based methods for shop floor control – both books are listed in the Further Reading section.

Simple visual reorder-point system

The simplest reorder-point system is one where a card is pulled when a material inventory lowers to a visually marked reorder point. It is often confused with the kanban system. This is appropriate for commodities where the value of having your finger on the pulse of the consumption is not worth the trouble. It can be implemented by simple means such as a red marked line in a stack of boxes or a painted red line on a transparent barrel. A reorder card is attached to the shelf and not to the item. Pulling it and handing it to purchasing uncovers a sign saying "on order."

CONWIP

Hopp, Woodruff, and Spearman described a concept they called CONWIP, for *constant WIP*, which, at first sight, looks identical to tunnel kanbans and appear to implement the same logic.[7] There is, however, a difference. In the job-shop system of Figure 9.32, priorities serve to sequence jobs prior release but not once they are on the floor.

With CONWIPs, sequencing in every center is done through dispatch lists based on priorities. While FIFO is not particularly desirable from a strictly scheduling point of view, it helps control quality by making the process easier to trace, and it can be implemented by simple devices like gravity flow racks. Dispatch lists, by contrast, require a system to produce them and require random access storage for WIP. While capping the WIP can produce order of magnitude improvements in the shop's performance, using dispatch lists might possibly provide a few more percentage points, and it is not clear that it is worth the cost, in particular the loss of process traceability.

POLCA

Other approaches that have been proposed for low-volume-high-mix environments include Rajan Suri's Paired-cell Overlapping Loops of Cards with Authorization (POLCA).[8] Unlike kanban, POLCA does not require a minimum inventory in each workstation, which will be uneconomical in a high-mix setting. The basic concept of POLCA is to use cards to show the free capacity between two cells. The idea is that the card loops between these two cells; it travels with the product from the first cell and when processing is finished in the second cell the empty card loops back to signal free capacity to process the next product. The second cell can be part of a POLCA loop with a third cell, and so on.

Each of the manual methods described above (kanban, visual markings, CONWIP, and POLCA) works well in its domain of applicability, but, as the number of pull loops in a plant rises, so does the administrative burden of managing the flow of physical signals, as well as the risks of misrouting. As a consequence the full or partial automation of these tasks using information technology is often the most tempting solution.

7 Spearman, M., Woodruff, D., & Hopp, W. (1990). CONWIP: A pull alternative to Kanban. *International Journal of Production Research, 28*(5), 879-894.

8 Suri, R. (1998). *Quick response manufacturing: A company-wide approach to reducing lead times*. Productivity Press.

9.4.6. Shop floor control with visual management

One thing is to control the materials that flow to the work stations, another is to control the work as it happens. In manufacturing, the term andon refers to a system that notifies workers of a quality or processing problem. The Japanese word andon means "paper lantern." There are many ways to implement andon, the most common ones include tower lights and andon ropes. Andon is simply a signal at the shop floor that attention is needed.

Tower lights

Tower lights are now found in many factories but are mostly misused for lack of a clear understanding of their purpose. When multiple small assembly lines or cells are operating in parallel, one tower light for each cell, placed above the last station along the transportation aisle, allows managers and supervisors at a glance to see which lines are operating, which are idle, and which are down. Since an assembly line is either on or off as a unit, tower lights are not needed at every station.

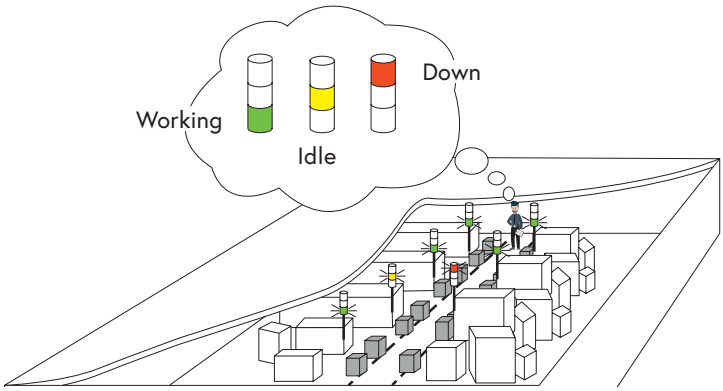


Figure 9.33 Effective use of tower lights

The rules for proper usage are as follows:

1. **Green, yellow, and red only.** The same three colors used in traffic lights should be the only ones allowed. There should be no white or blue light.
2. **One and only one color solidly lit at a time.** Blinking and combinations of light are not only unnecessary but harmful, since no one in management is able to tell the meaning of "solid white with blinking blue."
3. **Consistent meaning.** The meaning of the colors must be consistent throughout the plant, as well as be consistent with cultural conventions. Green may mean the line is running, yellow that it is idle, and red that it is down. If that is the convention, it must apply to every tower light in the plant. Green must not be used to indicate that a line is down or stopped, as this would be inconsistent with the cultural conventions established by traffic signals.

Green, yellow, and red tower lights are a simple and effective communication tool, which suppliers and users are conspiring to make complicated and unintelligible by adding more colors of lights, allowing combinations of lights with slow blinks, fast blinks, and no-blink patterns, and assigning to these combinations meanings that vary with each station (see Figure 9.34).

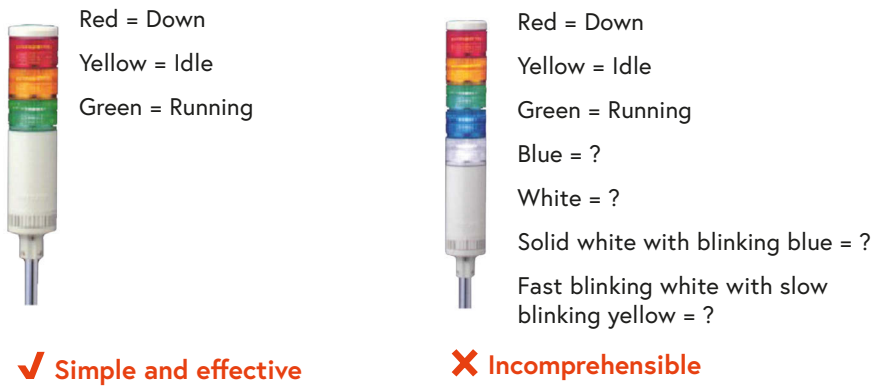


Figure 9.34 Good and bad design for tower lights

Andon rope

More elaborate communications require different devices. Large, moving assembly lines are fitted with stop ropes that allow assemblers at any station, and anywhere within that station, to stop the line. Pulling the rope is effective at keeping the line from producing defective units, but, by itself, does not help bring the line back up. For this purpose, Maintenance needs to know on which of the possibly 200 stations on the line to intervene.

This is done by a combination of visible and audible devices, as shown in Figure 9.35. The visible part is a panel displayed in a central location with lights showing the section of the line the stoppage originated in and the status of the response. The audible part is a section specific tune played on the public address system. For example, Gioachino Rossini's William Tell overture may point to a particular section of the trim line, and Edvard Grieg's Morning Mood to powertrain.

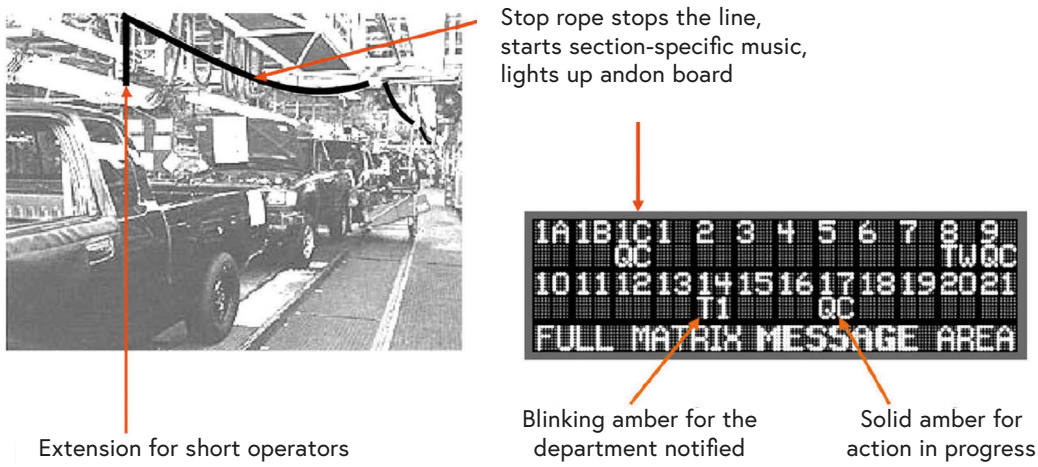


Figure 9.35 Stop rope and andon boards

Today, although such analog systems seems outdated, they remain useful for many factories. More progressive factories are implementing digital andon systems that call the attention of only relevant people. An example GKN Sinter Metals use of smartwatches, showcased in Box 9.5.

Box 9.5 **Smartwatches as digital andon in GKN**

GKN Sinter Metals uses smartwatches for shop floor control in factories in Italy and Germany. These factories make sintered metal parts by the millions for automobile industries, selling them for US$0.05 to US$1 per part. The parts are made by mixing metal powder, compacting it with pressure, and baking them in an oven.

In GKN, whenever a machine stops or is in need of replacing the carts for full products, shop floor personnel that "subscribe" to that machine get a message on their smartwatch that tells the machine's status. The person can accept or reject to do the task directly on the smartwatch. If accepted, the person carries out the task and reports back via the smartwatch. If rejected, the task will eventually be elevated to the next hierarchical level.



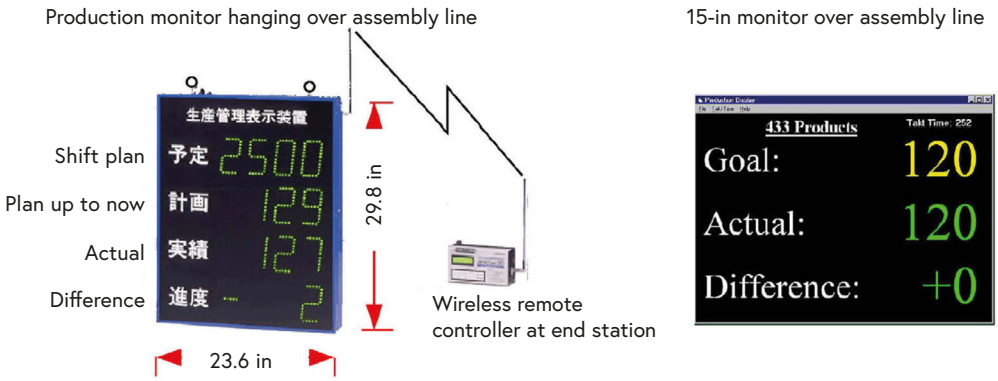
Figure 9.36 Smartwatches in GKN Sinter Metal's factory in Bonn, Germany

The smartwatch system has some distinct advantages over wall-mounted andon boards: First, personnel gets the message wherever they are, which allows them to be used more flexibly across the factory. Second, messages are only sent to relevant and qualified personnel. Third, performance can be tracked and measured at the individual level, which again can be used to make better use of everyone's time and balance workload across teams.

However, there are also many challenges and open questions regarding the use of smartwatches as digital andons. Economically, a critical concern is whether the large associated costs pay off in productivity increase. Technologically, a considerable digital infrastructure is needed to install and maintain the system. Organizationally, the system has to be managed, employees to be trained, and new work routines implemented. And, behaviorally, the smartwatches are problematic in terms of privacy and for some factory workers difficult to operate. Time will tell if smartwatches will take over the job of old-fashioned andon boards.

Counters and production monitors

While moving conveyors enforce the takt time, they are not present at every assembly operation. In that case, the assemblers need other means of knowing whether they are working too slow, too fast, or at just the right pace. Production monitors such as the ones shown in Figure 9.37 serve this purpose in many assembly cells and short lines.



a. Production monitor commercially available in Japan

b. Production monitor internally developed with obsolete PCs

Figure 9.37 Production monitors

The displays are located above the last station of the line, and typically show cumulative numbers from the start of the shift, including how many units should have been completed so far, how many actually have been, and the difference between the two.

One key engineering issue is how the production monitor is told that one more unit of product has been completed. Asking the assembler to press a button is simple but adds work and is error-prone. Arranging for finished units to pass by a photo-eye on the way out is a better way.

Monitors such as the one from Herutu shown in Figure 9.37a are available from several vendors under names like "industrial scoreboard," "production scoreboard," or "production pace timers." Rather than using wireless communication, many are connected by a wired network to a programmable logic controller (PLC).

9.5. Enterprise Resource Planning

For decades, the market for manufacturing software has been dominated by generations of products, called by a succession of names, all ending in "planning":

- MRP – Materials Requirement Planning (1960s)
- Closed-loop MRP – (1970s)
- MRP-II – Manufacturing Resource Planning (1980s)
- ERP – Enterprise Resource Planning (1990s–)

The MRP planning model from the 1960s making calculations from simplistic models is still the planning engine in today's ERP systems. As of 2020, per Statista, worldwide sales of ERP systems were $94 billion, compared to $40 billion for industrial robots, $18 billion for MES, and $11 billion for PLCs. Most ERP systems are now used *outside* of manufacturing and planning is only a minor part of them.

9.5.1. From MRP to ERP

Each generation added more functions to support planning, including master data management and commercial transaction processing, and these "peripheral" functions have overtaken planning as the main reason manufacturers use these systems. Figure 9.38 shows the external context of the original MRP.

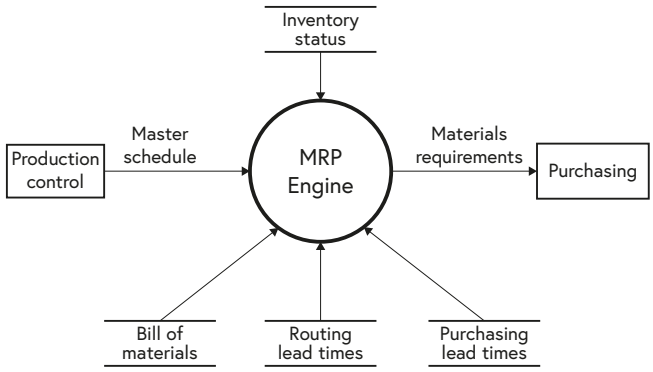


Figure 9.38 Materials Requirements Planning (MRP) in the 1960s

The original concept of MRP is of interest because it is still embedded in today's ERP systems. Wight and Orlicky introduced the concept of computing the "dependent demand" for components from the "independent demand" for finished goods as an improvement over methods like the two-bin or reorder point, which make no use of the bills of materials. A manufacturer should not have to reorder based on inventory levels alone when it is possible to calculate exactly the quantities required in order to satisfy the demand for finished goods.

One immediate concern with the picture is the quality of the input data. How can we make sure we feed MRP with accurate inventory data, bills of materials that incorporate the latest engineering changes, and lead times – both internal and external – that are realistic?

The answer is "closed-loop MRP" shown in Figure 9.39. In black is what was already in plain MRP. Closed-loop MRP added new recipients for the outputs of the system, and new linkages, some of which embed dubious business logic, in particular the use of dispatch lists to drive shop floor operations. In most plants, these lists are not actionable, for lack of the resources needed for the top tasks. Also, tracking shop floor activity – "shop floor control" – requires extensive online transaction processing. Shop floor transactions are often not entered in real time by operators, but at the end of the shift or the day by a clerk, from data handwritten on travelers, causing the inventory status to be off by several hours on a routine basis.

The capacity model supported in the closed-loop MRP system is limited to only one parameter for each resource type: resource time per unit, or units/hour on the resource. It contains no consideration of the following:

- Differences between machines of the same type.
- Allocation of work among machines with overlapping capabilities.
- Setup and production run structure.
- Machine load size.

This lack of details limits the validity of this black box or pipeline model to a level of aggregation at which operational details are blurred. It works for sales and operations planning on a monthly level, but not for shop scheduling on an hourly basis.

The diagram also shows inventory status as being maintained by tracking shop floor activity. In practice, this needs to be supplemented by cycle counting, that is, manually checking inventory status in set intervals, resulting in corrections to the inventory status database that are not shown here.

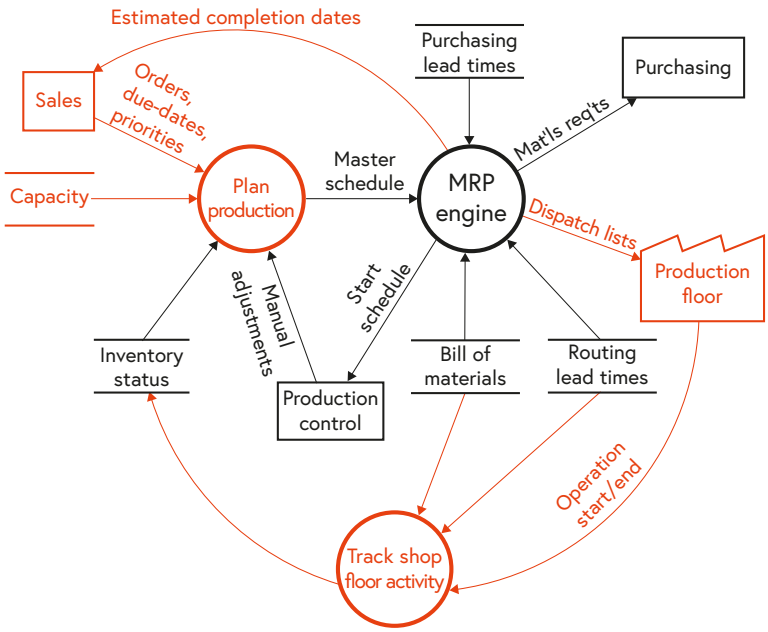


Figure 9.39 Closed-loop MRP in the 1970s

A decade later, in the 1980s, the meaning of the acronym "MRP" changed to "Manufacturing Resource Planning," or MRP-II for short, suggesting a broader focus than just materials. This next level of scope creep is outlined in Figure 9.40. As in Figure 9.39, the added functions are in red.

The automation of order entry was the most successful new feature. The customer database contains all the contract terms and conditions, customer-specific discount schedules, which makes for prompt, error-free quotes and invoices. Although this was not commonplace in the 1980s when MRP II emerged, direct order entry by customers through Electronic Data Interchange (EDI) is today.

Also developed since the 1980s is the practice of vendor-managed inventory (VMI), whereby customers grant suppliers access to their inventory database and make them responsible for issuing their own replenishment orders. Within lean manufacturing, pull systems were originally developed before the technology for EDI or VMI existed. It does not mean that EDI or VMI are in any way incompatible with pull systems.

Also added in MRP II is the ability to extract data from the MRP II database for use in other applications. Critical to the value of this feature is the extent to which the data can be trusted. Commercial data about orders and shipments are often accurate; equipment capacity, rarely.

Inventory accuracy cannot be taken for granted. Its achievement depends on a combination of management policies that drive the quantities of materials stored, visual systems to make items and quantities on-hand easy to see, and computer-based warehouse management systems (WMS) that, while not part of MRP II, have been successful in many companies.

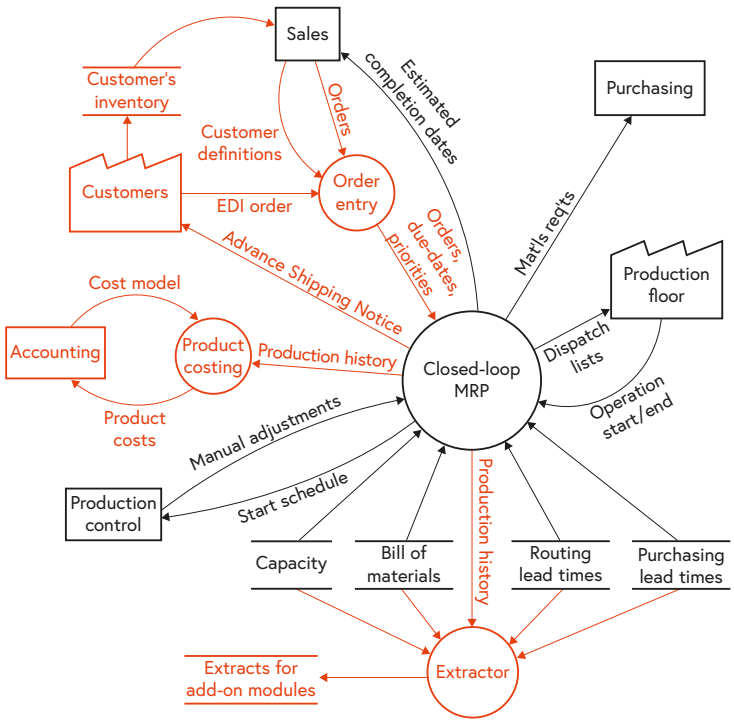


Figure 9.40 Manufacturing Resource Planning (MRP-II) in the 1980s

The last major functional expansion came in the late 1990s with "Enterprise Resource Planning," or ERP. Twenty years later, this name is still current. The focus of change since 2000 has primarily been the migration to cloud-based systems.

ERP adds another set of functions that go beyond manufacturing. The selling point is integration. The question is whether a single software supplier can have the application expertise needed to provide the best systems in domains as different in technical and managerial substance as maintenance and human resource management.

PLM (Product Lifecycle Management) is the integration of CAD with PDM (Product Data Management), which ensures the consistency and currency of the BOM data used in purchasing, engineering, and manufacturing.

The approach to external interfaces is richer than in previous generations. Extracts are of course still possible, but data warehousing makes them easier to generate for analytics, and real-time interfaces through middleware are available for bolt-on modules involving transaction processing, such as warehouse management. Using the same conventions, Figure 9.41 shows common functions added to MRP-II in ERP.

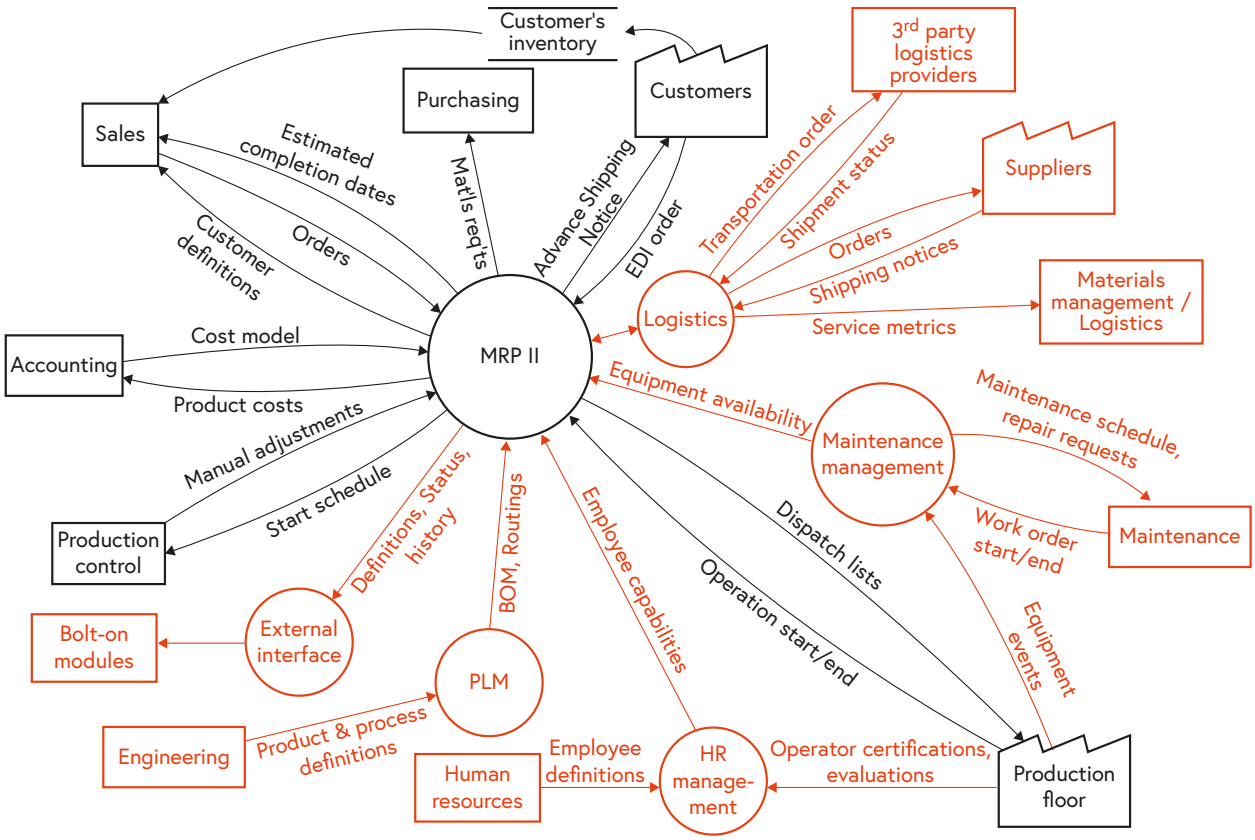


Figure 9.41 Enterprise Resource Planning (ERP) since the late 1990s

9.5.2. **Limitations of ERP**

Since its beginning as MRP, ERP has been a top-down approach, which consists of first dealing with planning at the highest level of the company, and cascading down to the production floor. It is the opposite of the other methods described in this chapter, which start with the scheduling of tasks, at the level where the instructions translate into actions on the materials, and then proceed to higher-level production and planning.

The algorithms at the heart of today's ERP are the same as in MRP 50 years ago. Figure 9.38 to Figure 9.41 show this accumulation on the same brittle foundation. MRP uses summaries, shortcuts, and approximations often based on tenuous or unrealistic assumptions, sometimes forcing concepts on reality to allow the use of a tool. This means that MRP fails at the *pragmatic level*.

A computer system serving operations can be described at three levels:

- **Syntax.** For example, references are identified by a series of five uppercase letters and numbers and dates are in day / month / year format.
- **Semantics.** These are the rules for using concepts. If an operation has a standard lead time of 2 days and requires 4 screws, you need to plan the availability of the 4 screws 2 days in advance.
- **Pragmatics.** This is the extent to which the abstractions in the system match reality. If all parts are available, the operation takes 10 minutes, not 2 days.

When we implement MRP, we assign to each operation a lead time that includes both the waiting time and the duration of the operation and is treated as a constant. We then add these lead times for successive operations and we measure the performance of production groups on their ability to hold them.

In reality, lead times vary and, to protect themselves, the first-line managers lobby for planning lead times set at extreme values that they can easily meet. They are usually successful. When we add up successive lead times, however, we add *extreme values*, which is logically incorrect because the extreme value of a sum is *not* the sum of the extremes values of its terms.

If your process has 100 operations, each with lead times that vary on a bell curve with a mean time of 60 minutes and a standard deviation of 10 minutes. Then the 99th percentile of the duration is $60+2.326\times10 = 83.26$ minutes. If you add the 99th percentiles for 100 operations, you get 8,326 minutes = 139 hours.

As the means add up, the mean time through 100 operations is 100 hours = 6,000 minutes. The standard deviation of the sum, on the other hand, is $\sqrt{100}\times10 = 100$ minutes, and the corresponding 99th percentile at $6{,}000 + 2.326\times100 = 6{,}233$ minutes = 104 hours. Adding the 99th percentiles has overstated the 99th percentile of the total lead time by 34%, and the more variability there is, the more pronounced this effect.

To make matters worse, ERP systems sequence the work on the floor by critical ratio – that is, the ratio of time available until due at the end of the process to the total lead time of the remaining operations. Sequencing by critical ratio ensures that only tasks that are on time or late are executed, with the result that, on average, all tasks are late.

Workarounds

The users of ERP systems commonly work around the limitations of their system's planning capabilities by using Excel or other electronic spreadsheets, which causes problems like the omission of a third of the items in an order due to an incomplete copy/paste. When selecting a data table in a spreadsheet with the cursor, the operator may mistakenly leave out the last lines. Other mishaps include products with no sales but positive revenue, and incorrect formulas.

Even in the absence of errors, the proliferation of copies of extracts from the system poses problems of update consistency and data security. Meeting participants have files on their laptops with the same names but containing data extracted at different dates and modified without coordination.

The use of spreadsheets or other workarounds in daily planning is clearly a dysfunction but so deeply rooted in the culture of factories that few practitioners recognize it as the root cause of problems they are experiencing.

As discussed in Chapter 8, there are better ways of sharing data between applications in an organized and controlled way, for example by means of a data warehouse or lake for historical data and an in-memory database for the current state.

The DDMRP update

DDMRP is an effort by Chad Smith and Carol Ptak[9] to address the dysfunctions of the MRP component of ERP systems. The "DD" in DDMRP stands for "demand-driven," which is odd because MRP is already supposed to be demand-driven. Unlike most approaches to production planning or scheduling, DDMRP is not incarnated in a software product but in a functional specification. Its creators, Smith and Ptak certify ERP systems and retrofits to ERP systems as supporting DDMRP.

In their own words, DDMRP has five components:

- **Strategic inventory positioning.** Buffers are needed whenever items have order fulfillment lead times from suppliers that are longer than required by the customer, as in when you need delivery within 24 hours of an item the supplier needs to start making 4 months in advance.
- **Setting buffer profiles and levels.** The next step is to determine the amount of protection at each buffer points.
- **Dynamic adjustments** are changes to that level of protection based on operating parameters, market changes, or anticipated future events.
- **Planning** is the process of generating purchase orders, manufacturing orders, and stock transfer orders.
- **Execution** is the process by which a DDMRP system manages open supply orders.

9 Ptak, C. A. & Smith, C. (2019). *Demand Driven Material Requirements Planning (DDMRP): Version 3*. Industrial Press, Incorporated.

Summary

The most essential points in this chapter were as follows:

1. The length of planning horizons depends on the business. Industries with fast clockspeed have short horizons (for example, fashion products and smartphones). Industries with slow clockspeed can have long planning horizons (for example, cruise ships and raw metals).
2. Sales and forecasts are used to generate the Master Production Schedule, which again is broken down to weekly and daily production schedules.
3. A balanced schedule helps smoothe the demand on shop floor resources and reduce inventory build-up.
4. Autonomous shop floor control uses visual cues and signals to communicate the need for an action, without the interference of a central IT system.
5. The two-bin system is the simplest reorder level system and is effective for small, inexpensive parts used in large volumes.
6. Kanban, developed by Toyota, is the most widely known card-based control system. While implementation differs between companies, we can separate four generic kanban systems:
   - Common kanban, which uses one card for both transportation and production
   - Hybrid system, which separates move kanban and production kanban
   - Tunnel kanban
   - Supplier kanban
7. Visual shop floor control systems, such as tower lights, andon ropes, and counting boards, help keep the pulse on running production and intervene when needed.
8. ERP developed from MRP, closed-loop MRP, and MRPII. MRP is still the planning engine in ERP systems.

Test yourself

1. What are typical short-term, medium-term, and long-term planning horizons for the following products:
   - Airplanes
   - Cars
   - Smartphones
2. What does the Master Production Schedule tell and what does it not tell?
3. What is the purpose of balancing the workload on the shop floor?
4. What is an "every-product-every" strategy and for what type of manufacturing can it be beneficial?

5 For what type of parts are autonomous shop floor control systems commonly applicable?
6 How do the two-bin system, the common kanban, and hybrid kanban system work?
7 What is the main difference between kanban and POLCA or CONWIP?
8 What is andon?
9 What are the key calculations in MRP?
10 What is the role of planning in ERP?

Further reading

Galsworth, G. D. (2011). *Work that makes sense: Operator-led visuality. Creating and sustaining visuality on the value add level*. Visual-Lean Enterprise Press.

Roser, C. (2021). *All about pull production: Designing, implementing, and maintaining kanban, CONWIP, and other pull systems in lean production*. Allaboutlean.com Publishing.

Sugimori, Y., Kusunoki, K., Cho, F., & Uchikawa, S. (1977). Toyota Production System and kanban system materialization of just-in-time and respect-for-human system. *International Journal of Production Research, 15*(6), 553–564.

Thurer, M., Stevenson, M., & Protzman, C. (2016). *Card-based control systems for a lean work design: The fundamentals of kanban. ConWIP, POLCA, and COBACABANA*. CRC Press.

Vollmann, T. E. (2005). *Manufacturing planning and control for supply chain management*. McGraw-Hill.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

10 Master Data Management

Data is to manufacturing what nerves are to muscles. Customer orders, delivery times, product and process specs, critical dimensions, machine status, workforce availability and skills, performance measures – these are all data without which a factory cannot operate. Within this context, this chapter focuses on *master data*, which we define and distinguish from other types of data in terms of relationship to time.

Master data is timeless, in that it changes only as the result of management decisions. Master data includes the product nomenclature, bills of materials, routings, assembly instructions, supply chain management procedures, customer data, and more. It is foundational for all manufacturing activities. On a daily basis, master data is the law. Occasionally, however, the law changes, as managers retire obsolete products and introduce new ones, engineers modify processes, and purchasing qualifies new suppliers. As such changes have far-reaching consequences they cannot be made lightly or casually.

How manufacturers manage and maintain their master data separates winners from losers, and its importance is only increasing in relevance with the ongoing digitalization of manufacturing. This chapter gives an overview of the structure of manufacturing master data and the challenges of updating it both prudently and promptly.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Differentiate master data from other data.
2. Read, generate, and maintain master data.
3. Understand the "key" approach as an alternative to "smart" product identifiers.
4. Explain the importance of nomenclature.
5. Generate a bill of materials.
6. Explain the relationship between bill of materials and routing.

DOI:10.4324/9781351110310-13

10.1. Introduction to master data

Managing master data well is of the utmost importance in manufacturing. It was a challenge in paper form; it is a different one in electronic form. To maintain data integrity, all master data needs a single-source-of-reference. Every class of master data should be stored in one place in an organization that is defined and agreed upon as the single source of truth – even if the data is duplicated elsewhere. This place can differ for different types of master data. For example, signed contracts can be stored in a central and physical library, customer contact data are usually stored in the ERP system, and product drawings can be stored in a PLM database.

10.1.1. What is master data?

In manufacturing, the master data include both lists of *items* and *relationships*. The kinds of data treated as master data varies among industries and even from one manufacturer to another within the same industry. Some of the most common *items* include the following:

1. **Nomenclature.** The unique IDs given to items like products, materials, WIP items, machines, methods, employees, suppliers, customers, etc.
2. **Operations.** Specifications of outcomes, work instructions, and process programs for automatic machines.
3. **Equipment.** The machines, tools, vehicles, and facility systems in the plant.
4. **Stores.** All the areas used for storage of materials or tools, including the locations and characteristics of each storage location.
5. **Human Resources (HR).** Names, addresses, contact details, bank account, emergency contacts, employment contract, education, skills, etc.
6. **Suppliers.** Names, addresses, contact persons, contact numbers, agreements, contracts, etc.
7. **Customers.** Names, addresses, contact persons, contact numbers, agreements, contracts, etc.
8. **Alliance partners.** Businesses sometimes called "complementors" are neither suppliers nor customers but their products or services enhance the value of yours. Examples include software publishers for computer manufacturers and oil companies for car makers.
9. **Intellectual Property (IP).** The patents, trademarks, and licenses owned by the company.

The most important *relationships* include the following:

1. **Bills of materials (BOM).** The network of items being assembled into other items, split into different items, or binned into different categories of products.
2. **Routings.** The sequences of operations performed between nodes of the BOM.
3. Organization structure. Divisions, departments, job descriptions, and the people occupying the different positions.
4. **Supply chain information.** Which components are bought from which suppliers, which products are sold to which customers, and under what terms and conditions.

For each entry in the above lists, the master data include two layers:

1. An information model that specifies, for example, the method for generating product IDs or the way the details of operations are attached to routings.
2. The population of this model; for example, the actual list of products made in the plant and the routings by which they are made.

The master data provides the explicit rules the organization works by. In addition, there are undocumented, implicit rules usually described as "the way we work around here," also known as the organization's culture. Note that the culture often overrides the explicit rules, with employees following habits, customs, or their personal ethics, all of which may break the rules. Culture, however, is not the subject of this chapter. The point of mentioning it here is to put the master data in perspective. A factory cannot function without master data, but factories with the same master data can operate and perform differently based on different cultures.

With respect to time, the other manufacturing data fall into the following 3 categories:

1. **Plans and schedules** are about the future, and are altered by the passage of time, whether successfully executed or not.
2. **Status** is the present. It changes as the result of events like operation starts, operation completions, machine failures, materials passing their expiration dates, or accidents.
3. **History** is the past. It never changes but keeps accruing.

10.1.2. Problems with master data

Unfortunately, managing master data is not as easy as it may seem. The problems with master data are numerous:

- **Errors.** Errors in master data cause friction in operations and defects in products. It also leads to confusion in the organization and can be disastrous for a company.
- **A simplistic information model.** For example, operations on a product commonly have time fields for "setup" and "run" when in fact the setup time for this operation depends on both the product you are changing from as well as the product you are changing to. A more comprehensive structure to hold this data for this operation would be a setup matrix with a different time for each from/to pair of products. Sometimes, this matrix is symmetric and sometimes not. Sometimes, it can be reduced to setup times between families of products. And sometimes, setup times depend not only on the last product made but on the entire sequence of the products previously made. It is also dependent on the experience of the operators.
- **Obsolete concepts.** In evolving forms, manufacturing master data has existed since the Industrial Revolution. Companies have developed and maintained them over time with the available information technology. Until the late 1990s, most companies kept the bulk of their master data on paper, organized, and filed accordingly. As of 2020, they are in databases, and some of the methods that served companies well in the past have become counterproductive.
- **Item proliferation.** Whether for purchased components or cutting tools, manufacturers' item lists tend to grow longer than necessary as engineers design new parts, and choose new components or tools instead of reusing existing ones. When Aisin Seiki's Kariya plant, Toyota's single source of proportioning valves in Japan, burned in February 1997, Toyota realized that it was using 200 varieties of these valves. Restoring production in an emergency for 200 varieties was more of a challenge than for the roughly 30 that would have met the company's needs.
- **Cumbersome change management.** Management decides when to change master data. As these changes have far-reaching consequences, they are subject to an approval process that companies struggle to keep both thorough and prompt. When you routinely have 500 engineering change orders (ECOs) pending with a 4-month lead time, the system is dysfunctional. In emergencies, managers will authorize engineers to work around the system, which may solve a problem in the short-term but come back to haunt them later. In one case, engineers dispatched to a supplier changes in the design of a purchased component by manually annotating drawings, which solved a

problem. Two years later, however, the managers decided to terminate the supplier, and discovered that the only documents containing the correct design for the component were the drawings in the hands of the supplier.

10.1.3. Master Data Management software

To "solve" some of the problems listed above, many IT companies offer Master Data Management (MDM) software. This software promises to be a single-source for master data in execution, with the ability to revise it in an expeditious but controlled manner. Figure 10.1 is a simple example of the concept, as actual systems involve more participants.

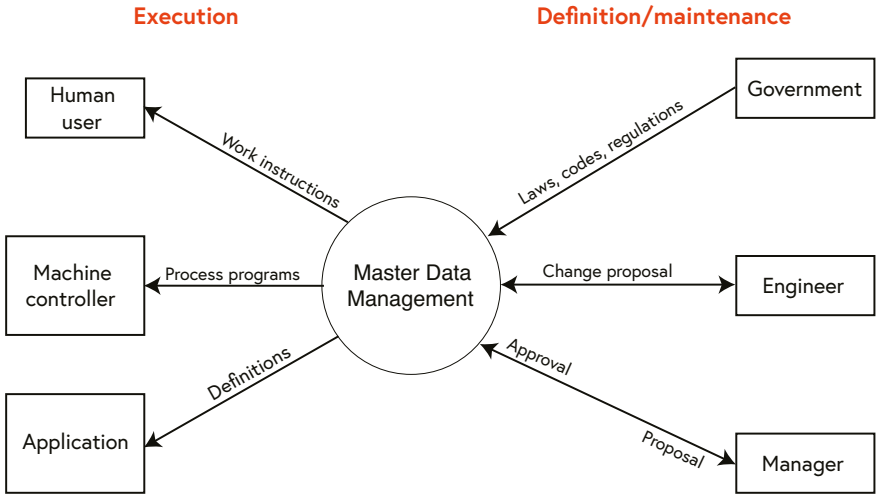


Figure 10.1 MDM software aims to be the single-source-of-truth for master data

Managers of manufacturing companies have been slow to see the value of such tools for their master data. In fact, many only started implementing them because it was a requirement for ISO 9001 certification. Today, the market for MDM software is booming as many companies try to sort out their data in their "digital transformation" journeys. In the early 2020s, the top MDM players include, among many others, SAP, Oracle, IBM, SAS Institute, and Microsoft. As usual, each vendor offers to solve all master data problems, which is not a reasonable expectation.

The question of what MDM solution to choose should not be taken lightly. The management of master data is not something manufacturers should outsource to vendors. Master data is critical to operations. It must be shared and readily available to parties with a need to know, yet protected from competitors. Process definitions in individual employees' Excel spreadsheets, for example, meet none of these requirements, but is still common practice in many firms.

10.2. **Nomenclature**

Every entity involved with a factory needs a unique ID, to which its other characteristics can be attached. This includes products, materials, purchased components, processes, operations, machines, workstations, tools lines, storage locations, suppliers, etc.

What you call parts, products, and resources determines how easy they are to retrieve in daily operations and to analyze for planning or control. This includes the technical, commercial, and administrative data attached to their names. Nomenclature matters.

10.2.1. **"Smart" product identifiers**

Manufacturing, worldwide, is overwhelmingly dominated by part numbering systems that are called "smart" because they embed data about materials, dimensions, number of holes, or other characteristics into part numbers, as in CEP-134M10FB-TT1234\R4, where each group of characters has an encoded meaning, for example CEP for "capacitor" and R4 for "revision 4." Figure 10.2 shows another example of a typical product nomenclature. The product is a microcellular polyurethane sealing product manufactured by the Rogers Corporation.

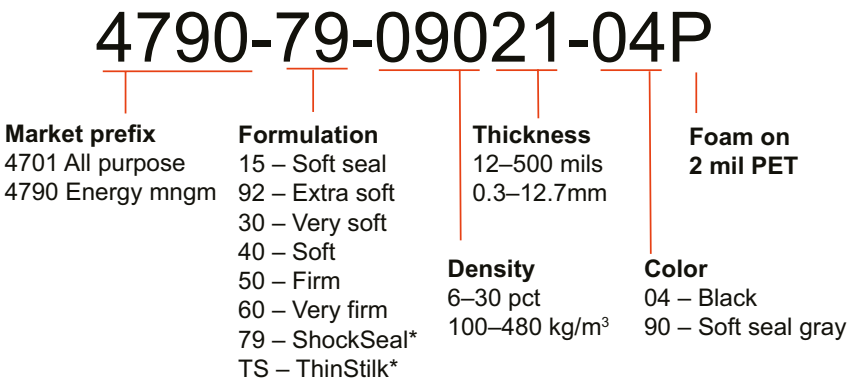


Figure 10.2 Example of a product nomenclature for a sealing product

"Smart" part numbers have grown such deep roots in manufacturing operations that few managers or engineers are even aware of the unnecessary costs they generate in training and routine decoding, and of the obstacles they place in the way of manufacturing data mining.

"Smart" part numbers give similar names to similar parts, which makes them more difficult to tell apart during picking and increases the risk of picking errors. Hergé's Tintin stories feature the two twin, dim-witted detectives named Thompson and Thomson shown in Figure 10.3. They only differ by the shapes of their moustaches, and their nearly identical names don't help. An example in manufacturing are two nearly identical screws and their nearly identical part numbers. One of the screws is 1/8 inches shorter than the other, which, with "smart" part numbers, results in only the last digit being different.

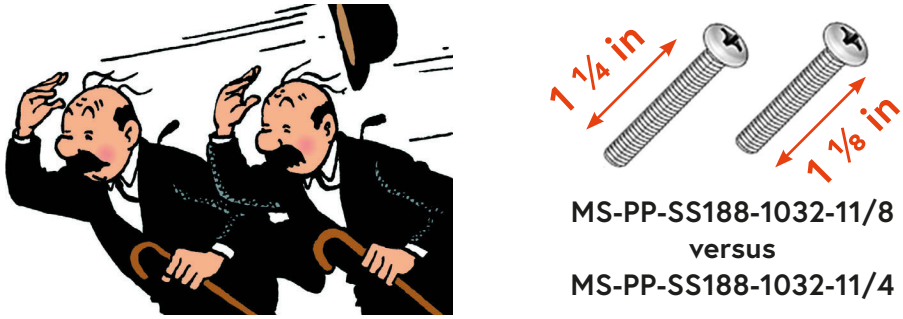


Figure 10.3 Thompson and Thomson by Hergé and two almost identical screws

The "smart" nomenclature systems are a 100-year old legacy of the paper-and-pencil age, and obsolete in the age of databases. The better alternative is to have part numbers that are just unique IDs, or *primary keys* in database terms, with all the data attached in the form of property lists in plain text. We refer to as it the "key" approach.

Box 10.1 **Origins of "smart" numbering systems**

Smart numbering systems were introduced in libraries in the 19th century. In the 2011 movie *J. Edgar*, Clint Eastwood highlights the role future FBI director J. Edgar Hoover played in the development of the US Library of Congress cataloging system, in which "BX 378.5 .M38 1968" breaks down as follows:

- BX = Books on Religion
- 378.5 M38 = Shelf address
- 1968 = Year of publication

In the 1910s, this numbering system made sense because it allowed trained users to locate books quickly. A librarian or book lender could search the card catalog by author, title, or subject, and then locate the book by its call number, as shown in Figure 10.4.

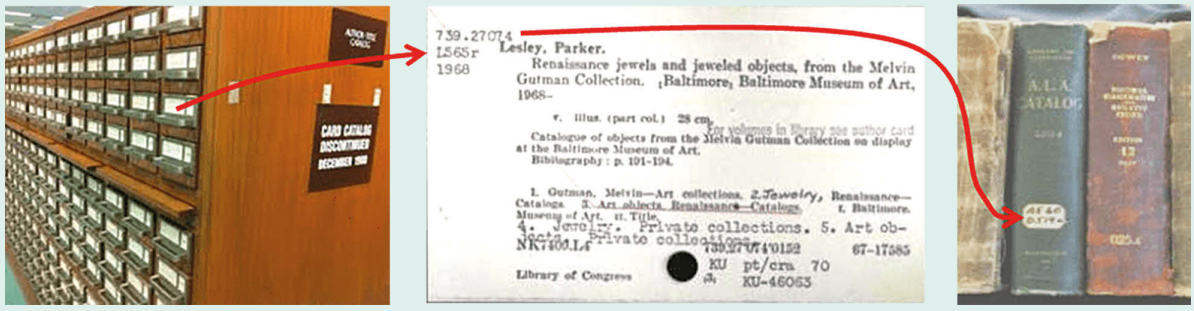


Figure 10.4 Using a library card catalog

10.2.2. The "key" approach

The "key" approach is not in any way less functional than "smart" numbers and it has other benefits:

1. It reduces training costs.
2. It is not affected when product information changes.
3. Part numbers are short, and therefore easier for humans to work with.
4. Part numbers are distinctive, even for similar parts, thus reducing picking errors.
5. Nomenclature errors are more easily detected and corrected.
6. Property lists in plain text on labels are easier to read.
7. Property lists support data mining.
8. Product catalogs from different companies can be merged without loss of information.
9. Company private information is not unintentionally disclosed through part numbers.

The case for the key approach is compelling, and the technology to use it available. But many manufacturing professionals active today have been educated in the old method, and taught that it was smart. Of course, people who have spent years mastering a complicated, antiquated system are always strongly attached to it and oppose its abandonment, as it would make their hard-won expertise irrelevant.

Table 10.1 shows an example of the key approach to an item and the characteristics that are usually embedded in a "smart" part number.

Table 10.1 The key approach to part numbers, in table form

Number	Item number	Property name	Property value
1	13AG5	Revision number	5
2	13AG5	Prior item number	D4E23
3	13AG5	Material	Stainless steel 316
4	13AG5	Category	Cylinder
5	13AG5	Diameter	in
6	13AG5	Height	in
7	13AG5	Volume category	A
8	13AG5	Warehouse location	18-5-2

The item is uniquely identified by the key "13AG5," and the nomenclature database has a record for each property of each item, in the form of a name-value pair. There are different views of what an item is. In some cases, the same key is used for all revisions, and you tell the revisions apart by their number. In this example, each revision has its own key, and contains the key of the prior revision in its property list, so that you can trace back the history of the part through multiple revisions.

With the key approach, all the properties can be changed without affecting the key, which does not work for "smart" numbers, and the complete history of production volumes, deliveries, and quality problems can be retrieved with the key. Neither the property names nor the values contain any abbreviations or codes. They are all in plain text. In particular, you may notice that, where applicable, property values include units, while "smart" part numbers usually don't. This is an important detail in multinational companies where metric and imperial measurements coexist. There is no need to store these

characteristics in short fields. You can use as many characters as you need. Thirty years ago, data storage space was expensive; now it is not.

Different items can have different property lists. Metal tubes have a length and a diameter; paint cans, expiration dates. The property values can also change over time without the item itself undergoing any change that would justify a new part number. For example, if the item is a product that undergoes a drop in demand, its volume category may be downgraded from A to B, and it may then no longer have a dedicated warehouse location.

Unless the users of part numbers, in materials handling, production, engineering, production control, purchasing, or sales, are able to retrieve the information embedded in a "smart" part number, embedding it is a waste of time. With "smart" numbers, hundreds or thousands of people would have to be trained on issues like the meaning of the field in characters 5 to 8 of a part number. In reality, it doesn't happen. Outside the department in charge of master data, very few have a clue. When asked what a part number means, they have to painstakingly decode it field by field from a spec. With the key approach and characteristics listed in name-value pairs in plain text, the information is accessible with no training, which is why it is used on e-commerce sites like Amazon.com. Each product's unique ID is Amazon's ASIN, which may designate anything from a book to a waffle iron. The product details are given in a list of name-value pairs, and this information is immediately obvious to a first-time visitor.

"Smart" part numbers are not only obscure but preposterously long, and therefore practically impossible for people to remember. No matter how large a field an ERP system or MES provides for part numbers, some company's "smart" numbering system will exceed it. In the key approach, the part number's one and only job is unique identification within the scope of the system.

There may be another object by the same name at the other end of the world, but not within the plant or the company. A sequence of just 5 uppercase letters and digits is mercifully short and easy to remember. It provides (10+26)5= 60.5 million possible IDs, and that is enough to meet the needs of even a high-mix production organization for a long time. Some prefer to use digits only, but it reduces the name space to only 100,000 IDs. Allowing lowercase letters would increase it to 916 million, but few people would reliably make the difference between "13AG5" and "13aG5."

Manufacturing professionals hesitate to remove "smart" numbers because the identifiers contain information that trained operators can use to separate parts from each other, as in the example above. The key approach, however, does not hide information from the human operator; it only separates information from identification. As shown in Figure 10.5, all the data in Figure 10.6, or a relevant selection from it, can be printed on a label.

13AG5	Revision **5**	Material **Stainless Steel**
Category **Cylinder**	Diameter **1.25in**	Height **25.00in**
Volume category **A**	Warehouse location **18-5-2**	Box quantity **50 pcs**

Figure 10.5 Box label with characteristics in plain text

Note also that all the characteristics on this label are in plain text. The best, cheapest, and least error-prone standard is using whole words from the native language of the workforce. If a bolt is called "Bolt," no training is needed to recognize it. If you call it "256," you need a dictionary. Human readability is, of course, not the only concern in daily operations. You also need to use barcodes, QR-codes, or RFID tags for machine readability.

Support for data mining

The most convenient input to manufacturing data mining is a table in the dimensional model. In the special vocabulary of dimensional modeling you have dimensions and facts. Some columns contain the dimensions used for filtering and aggregating. Other columns contain facts that are measured variables or observed attributes for which you want to identify patterns. All the characteristics embedded in a "smart" part number are dimensions. You want to be able to analyze quality for parts made from a given raw material, or within a size class, etc.

Retrieving this reference data by parsing "smart" part numbers is a major effort, which has to be repeated with each naming convention. With the key approach, on the other hand, turning the property list of Table 10.1 into Figure 10.6 is just a matter of using a crosstab query in Access or a pivot table in Excel.

Item number	Category	Diameter	Height	Material	Revision number	Volume category	Warehouse location
13AG5	Cylinder	1.25in	25.00in	Stainless steel	D4E23	A	18-5-2

Figure 10.6 Item properties crosstab for data mining

Crosstab queries and pivot tables transform lists of IDs and name-value pairs into tables but not the other way around. In R, the tidyverse package has functions called pivot_wider and pivot_longer to convert in both directions, which you need to do, for example, when consolidating item catalogs from multiple legacy systems that come in the form of tables with different column structures. You cannot append tables with different structures but you can always append lists of IDs and name-value pairs.

The ability to go back and forth between these two representations is gaining in value as there are commercial database systems based on both. As of 2020, the dominant model is the relational database management system (DBMS), in which all the data is in tables. These are commonly referred to as "SQL databases," based on the SQL structured query language. Oracle, MySQL, SQLServer, and Access are all SQL databases.

Databases that store data in the form of lists of IDs and name-value pairs are called "NoSQL databases," an unfortunate terminology because defining these systems by what they are not creates too broad a category. In fact "NoSQL" is short for "Not only SQL." NoSQL database products include MongoDB and Couchbase, which are used by major web services.

Detection and correction of nomenclature errors

The only way to be sure a part number is read and written accurately is by reading it automatically from a bar code, DataMatrix, or an RFID tag, and by printing it from the master database. Typing, handwriting, copying and pasting, and even selecting from a long pull-down menu with similar entries is error-prone, particularly when performed with data that has been extracted from an ERP, MES, or PDM software system into an electronic spreadsheet.

Besides typos, part numbers may have interchanged characters, be truncated, or be misinterpreted by the destination software, as when, for example, Excel treats a part number like a number and lops off leading zeros. And the slightest error in a part number is enough to make the consolidation of data about it from multiple sources impossible.

The above is true no matter what part numbering system you use. The risk of error, however, is much higher with "smart" part numbers than with keys, because they are 2 to 3 times longer and more similar, so that a part number with a typo may point to a different, existing part rather than nothing.

Nomenclature during mergers and acquisitions (M&As)

The possibility that the company might merge or be acquired is usually not a consideration in the design of its nomenclature. However, considering how often such events occur and how far reaching their consequences, it should be. When companies merge, their catalogs of products and parts should too.

Merging information models, and nomenclature in particular, is always difficult and time-consuming. Mergers are mostly of unequal companies, and the larger company's system becomes the merged company's standard, even when the smaller company has a better one. In mergers of equals, this situation often goes unresolved for years, with both divisions' legacy systems remaining in operation, and possibly, their real-time content meeting through integration middleware and their histories in a data warehouse.

As illustrated in Box 10.2, merging the property lists from two companies systems is work in any case but is easier to do with property lists than with "smart" part numbers, where it requires you to develop a sophisticated parser. The properties in both systems are likely to both have different names and sets of values for the same data.

The name value pair (Material, Steel) in one system may be (Mat'l, STL) in the other, and the engineers need to build a dictionary to translate them. At first, you may translate the data only in the data warehouse, which is sufficient to jointly mine the data from both systems, but, eventually, you would want to standardize the property lists across the divisions.

When you do this with the key approach, you deal with issues of meaning that have to be addressed anyway as part of a merger. With "smart" part numbers, you also have to resolve format issues that are unrelated to meaning.

Box 10.2 **Product nomenclature during mergers or acquisitions**

To focus on what happens with part numbers, we consider two hypothetical mergers. In the first, both companies have "smart" part numbers; in the second, both use the key approach. Real situations, of course, have many other patterns.

Scenario 1: Merging two companies with "smart" product numbers

Companies usually do not have identical "smart" part numbers. For them to use different naming conventions is good for uniqueness after the mergers, because it makes it highly unlikely that they will have assigned the same name to different products or parts. As the software systems themselves usually do not constrain the naming conventions – except for the total length of part numbers – you can upload another company's part numbers, and end up with a catalog that commingles part numbers based on both conventions. But this adds further complexity for data mining, because, to retrieve the information embedded in part numbers, you now need to parse names generated with *two* different conventions, and to identify which convention each name is based on. The management of the acquiring company intends to standardize the part numbers on its own system, but implementation has a way of being postponed indefinitely.

Scenario 2: Merging two companies using the key approach

Two companies using the key approach have a higher risk of key conflict, especially if the keys are sequential auto-numbers, because both companies will have products with keys 1, 2, 3, etc. This is a problem that is usually not anticipated when systems are first implemented, but can be avoided by instead selecting names at random from a large name space. For example, if 20,000 keys are chosen with equal probabilities among the 78.3 billion possible combinations of 7 digits and uppercase letters, then 99.7% of the time, no two will be identical. You would still need to search for duplicates just in case, but you only have a 0.3% probability of finding one.

The European Union had this problem with ZIP codes, and resolved it with country code prefixes. Another option, if you have two products both named 4384 by different companies is to rename one of them 4384.1. Any change to keys is undesirable, because all references to that key then need to be updated. This can usually be managed within one system, but not within all the ad-hoc extracts and spreadsheets that users have generated.

Stop and Think 10.1
Nomenclature is important but hard to maintain. What do you think are the main reasons why companies struggle to establish nomenclature and keep it current?

Data security

The protection of proprietary information is the only legitimate reason to encode and encrypt data. During development, for example, the marketing names of products are not used for fear that they might leak to competitors. Once a product is made and sold, all units often bear serial numbers that must be encrypted lest they provide competitors with information about production volumes. Consider the following examples:

- The use of plain text in serial numbers on World War II German tanks allowed British analysts to accurately estimate the numbers produced from the serial numbers of the units that were captured or destroyed (Figure 10.7). The statistical inference problem is known as the German Tank Problem.
- The serial number of an iPad today could be "DKVG805HDFHY," from which Apple competitors can deduce nothing unless they know how to translate it back to a production sequence number. "Smart" part numbers do not provide much security, as anyone who cares can reverse-engineer them from actual products and labels.
- For example, on the Lands' End catalog, once you find out that Men's Regular Short Sleeve Jacquard Polo shirts are item 40600-5A63 in regular sizes, and item 40600-6A69 in large sizes, it doesn't take Sherlock Holmes to guess that the numbers preceding the dash identify the polo, and the alphanumeric characters following it, a size class.

A "smart" part number is like a house key that would carry information about which door it opens; a thief cannot infer your address just from your house key because this information is not on it. Likewise, in the key approach, a product ID by itself will reveal nothing to your competitors.



Figure 10.7 A German World War II tank on the battle field

Versions and continuity

To issue operator instructions, or support customers, you may be tempted to have a different ID for each version. If you do, however, for all intents and purposes, different versions of the same product are different products, which is not what you want when analyzing market trends.

The original Volkswagen Beetle went through many versions from its launch in 1938 to the last one in Mexico in 2003, but it is still thought of as the same model. Adding all the sales figures for all these years reveal that 21,529,464 were built, making it the world's longest running and most manufactured car of a single platform.



Figure 10.8 Volkswagen Beetle production in Wolfsburg, Germany, and in Pueblo, Mexcio (Credit: Volkswagen)

For this to be possible, you need a version-independent product name. Version must be an attribute attached to the name that you use to retrieve the version-dependent technical information, but the different versions remain linked by the product name for all other purposes.

If you buy components to assemble into industrial products, you may find it expedient to just use the naming conventions of your suppliers and customers. You don't care how they came up with these names but, as long as they are unique, they support your communications with them, and you don't have to worry about generating your own. It sounds like a good idea but it's not for the following reasons:

- First, even if you believe in single-sourcing, you may not have achieved it yet, and you have some components with multiple suppliers, using their own part numbers. If you just use your suppliers' part numbers, and have four suppliers for the same bracket, you will have four items in your database that you will have to identify as alternatives, when all other items in a bill of materials are used in conjunction. In addition, it will make the number of components appear larger than it really is, underestimate the frequency of use of your components, and won't help you collocate functionally equivalent items in your warehouse.
- Second, you have no guarantee that no two suppliers will ever use the same name for different parts. It is unlikely, but, to make sure it doesn't happen, you add a supplier prefix to the names, which ends up doing exactly what you wanted to avoid: embedding information in the names.

The bottom line is that having your own names (IDs) for components, subassemblies, and finished goods, with no embedded information, will serve you better in production planning and logistics/supply chain management than to try to come up with "smart" product names.

10.3. **Bills of materials**

Everyone involved with assembly agrees that bills of materials (BOMs) are at the core of their activity. They would unfortunately also admit that BOMs often have chronic accuracy issues, that the work-arounds to these inaccuracies impair the companies' ability to update and customize products, and that maintaining BOMs is resource intensive.

> **Definition**:
> A bill of materials (BOM) is a list of all the parts and their quantities needed to manufacture a complete product and its subassemblies.

10.3.1. **What is a BOM?**

Over the years, usage of the term BOM has expanded beyond recognition. Starting from a product breakdown into subassemblies and components, it grew to include drawings, machine programs, operator work instructions, supplier contracts, etc.

If you ask non-manufacturing persons what they think a BOM is, they usually think that it's a form of shopping list. A shopping list is, in fact, the special case of a flat BOM. It contains all the items that have to be bought from the outside with the quantities needed in order to make a product but no information about the product's breakdown into subassemblies inside.

A general BOM contains both. For human readability, the BOMs of an assembled product is most commonly visualized as an upside down tree where the product itself is the root, purchased materials and components are the leaves, and subassemblies the intermediate nodes, marked with the quantities of each needed to manufacture a unit of the parent node.

Figure 10.9 shows an example of a BOM tree for Product X consisting of three subassemblies A, B, and C. The numbers in parentheses indicate how many of the parts are needed. Hence, Product X consists of 1 A, 2 Bs, and 1 C. Subassembly B consists of two Ds and one E.

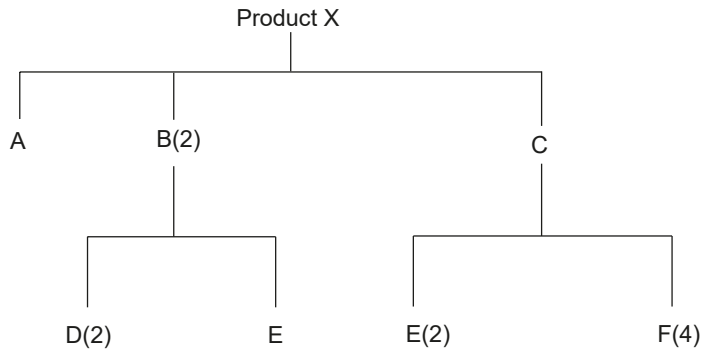


Figure 10.9 A BOM tree

The same item may appear in multiple nodes if it goes into multiple subassemblies, as for part E in Figure 10.9. This is common, for example with fasteners. If you print or display assembly instructions, it is

normal for the fastener to appear in the component list of every subassembly that uses it, with the quantity used. It should also be noted that the quantity used is not a function of the node but of the connection between nodes; 1 unit of E goes into a B, while 2 units go into a C. Therefore 1 belongs to the link between E and B, and 2 to the link between E and C.

As explained below, a BOM is actually not a tree but a more general directed graph, because the same leaf or subassembly can go into multiple other subassemblies or products. Not every directed graph is a valid BOM and we discuss below the constraints it must meet to be one. These constraints provide validation criteria to eliminate errors in actual BOM records.

In a BOM, the quantities are not attached to the vertices representing items but to the directed edges, and represent the amount of the origin item needed per unit of the destination item. For many purposes, it is often useful to attach other data to a BOM, like technical drawings and assembly instructions for subassemblies and final assembly, or supplier contracts and lead times for purchased items. For clarity, we limit our discussion here to the basic "goes-into," often called "Gozinto" structure that is common to all these different applications, from purchasing to engineering and manufacturing.

Note that printed documents, such as user guides, assembly instructions, and guarantee forms, that are shipped with the product and disposable packaging, including shrink-wrap, need to be in the BOM but pallets and returnable containers do not. Instead, they belong with the other reusable resources in the plant's operations, like fixtures and tools. The BOM should contain all and nothing but the materials that are consumed in the making of the product as shipped.

Box 10.3 **What if a product involves software?**

More and more products include software. How is that reflected in the BOM?

Consider the popular TipToi product of Ravensburger, a German toy manufacturer, as shown in Figure 10.10. TipToi is an interactive pen that children can use to explore physical books and boardgames. If bought as a "start-up package" with a book or game, the TipToi pen comes with one pre-installed program for that product. Other programs are easily installed by the customer on demand through a digital platform. How should the software for the startup package be reflected in the BOM?



The answer is, "not at all." Firmware items are not included in BOMs because they are not materials. They don't come in bottles or cases and don't require trucks. Embedding software into a product is an operation in a routing but it does not involve *withdrawing* units from stocks. Any needed hardware, such as a printed-circuit board or sensor components that is assembled in a physical product is of course part of the BOM.

Figure 10.10 TipToi from Ravensburger (Credit: Ravensburger AG)

Most companies do not make only one product. If we view each BOM as a tree, then, for the whole product mix, these trees add up to a forest. Each product variant has a BOM, but the same type of leaves or nodes can go into multiple product variants. If we view each one as a distinct network, we don't see this commonality. It can in fact be difficult or impossible to visualize when you are dealing with thousands of items.

> **Stop and Think 10.2**
> Can you make a complete list of materials needed to build Product X in Figure 10.9?

10.3.2. Invalid BOM records

Not every list of item pairs and quantities belongs in a BOM. The following, for example, are not allowed because they make no sense:

- An item is not allowed to go into itself. A cannot go into A.
- Loops are not allowed either because, indirectly, they lead to an item going into itself. You cannot have A go into B, B into C, and C back into A.
- The same pair of items cannot appear multiple times with different quantities. Item B either requires 10 A screws or 5, but it can't require both.

Graphically, these patterns appear as in Figure 10.11:

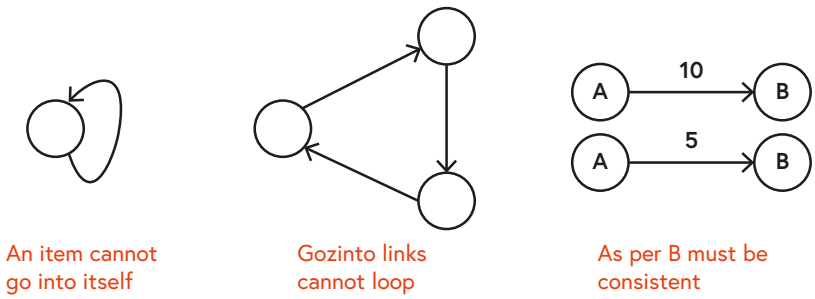


Figure 10.11 Wrong patterns in BOMs

Such patterns violate basic logic and make planning calculations fail but, unless BOM editing software flags them, it is easy to see how typos in IDs can make them creep into tables of tens of thousands of records maintained manually.

10.3.3. BOM accuracy

BOM accuracy is not easy to measure. Some authors separate completeness from accuracy. If an item is missing from a BOM, the BOM is incomplete; if it is listed with the wrong quantity, it is inaccurate. For most purposes, since both are errors, we can lump both under "inaccuracies." A straightforward measure of BOM accuracy would be the ratio of the number of products with BOMs that are free of errors at any level to the total number of products made by the company:

$$BOM\ accuracy1 = \frac{Number\ of\ products\ with\ error\text{-}free\ BOMs}{Total\ number\ of\ products}$$

In production, however, errors in the BOM of a product that is never made are harmless, while errors on a product made in high volume every day are crippling. We should expect BOMs to be maintained with greater care for the frequently made *runners* and *repeaters* and the errors concentrated among the sporadically made *strangers*. To reflect this, we could instead use the the ratio of the aggregate production volume for products with error-free BOMs to total volume:

$$BOM\ accuracy2 = \frac{Aggregate\ production\ volume\ for\ products\ with\ error\text{-}free\ BOMs}{Total\ volume}$$

A more important question is, however, why are BOM inaccuracies allowed to exist at all?

BOMs are data that is generated and maintained internally, as part of product design, and is not affected by variations in the business or technical environment. BOM entries are not influenced by market fluctuations or tool wear. They are decisions made by professionals, and BOM inaccuracies are discrepancies between these professionals' intent and the data they produce. Manufacturers can make smartphones, airplanes, or cars but they can't maintain a shopping list? Given that there is no technical reason, we have to look at human, social, and organizational causes.

Inaccuracies creep into BOMs as a result of the following:

1. Management pays attention to new products, not existing products.
2. Products are customized.
3. Products are designed and manufactured by different organizations.
4. Engineering changes are poorly managed.

For example, when a BOM is handed off by the Design team to Manufacturing, it makes a difference whether it is in the form of a copy of the design BOM or shared access in a database of master data. In the first case, only extreme vigilance will prevent the design and manufacturing BOMs from diverging; in the second case, they are technically prevented from doing so.

As a standalone project, fixing the BOMs is never-ending and tedious, and may even be futile when an improvement project on an assembly line or a change in purchasing policy renders obsolete BOMs that have just been audited. It is more effective to systematically include BOM audits as tasks within improvement projects. As you change the layout of an assembly line, or its subassembly structure, or the work content of different stations, shop floor observations will expose any errors in the BOMs and provide the opportunity to correct them. And the corrected BOMs will not be immediately obsolete.

10.3.4. **Who uses BOMs?**

Use cases are now the tool of choice for software developers to specify applications. They identify potential users and scenarios of the way they would use a tool. Then software developers build the tools to make all these scenarios actionable. Here we are attempting to do the same for the BOM that almost everyone working in a manufacturing plant uses in one way or another.

Perhaps Sales, Maintenance, and Human Resources don't need the BOM directly, but just about everybody else does, and for many different uses. The following are ways a few members of a manufacturing organization might describe how they actively use the BOMs:

- **Assembler**: "When assembling, I need to know which parts go into the product in which quantity at this operation."
- **Kit picker**: "To prepare kits for assembly operations, I need a pick list with quantities by item."
- **Design Engineer**: "When designing new products, I am being asked to reuse components from previous products."
- **Manufacturing Engineer**: "We have identified a better component for an assembly operation. I need to do the following:
  - Run a trial run with the new component in production.
  - Analyze the results of the trial.
  - If successful, submit an engineering change request order.
  - Secure approval of the change request.
  - Implement the change in production."
- **Quality Engineer**: "A customer has returned a defective unit. I need to trace the components it was built from."
- **Production Planner or Scheduler**: "I need to explode the product demand into the time-phased demand for purchased materials and components, and for internally made subassemblies."
- **Logistics/Materials Manager**: "I need to ensure that the right parts are available in the right quantities at the right time for assembly to execute its schedule."
- **Supply Chain Manager**: "I need to define a plan for every part (PFEP) we buy, including ordering method, frequency, and quantities, based on the production plans and schedules." Even when you use the method that is simplest for the customer – consignment – you use the BOM to calculate the quantities used in the finished goods to calculate how much to pay the supplier.
- **Accountant**: "I need to assess the financial impact of changes in terms and conditions with suppliers, based on the current and future workload and the bill of materials."

> **Stop and Think 10.3**
> LEGO toys usually include a build-instruction that lists all building blocks that are part of a build set. Is that a BOM?

Box 10.4 **BOM in ERP systems**

It is most common for firms to maintain their BOMs in their ERP system. It can be quite confusing to navigate between different ERP systems and understand the role and parameter setting for the BOM. Different ERP systems have somewhat different labels and solutions for doing so. Figure 10.12 shows the SAP interfaces for managing the material part of the BOM for a computer manufactured by Sony.

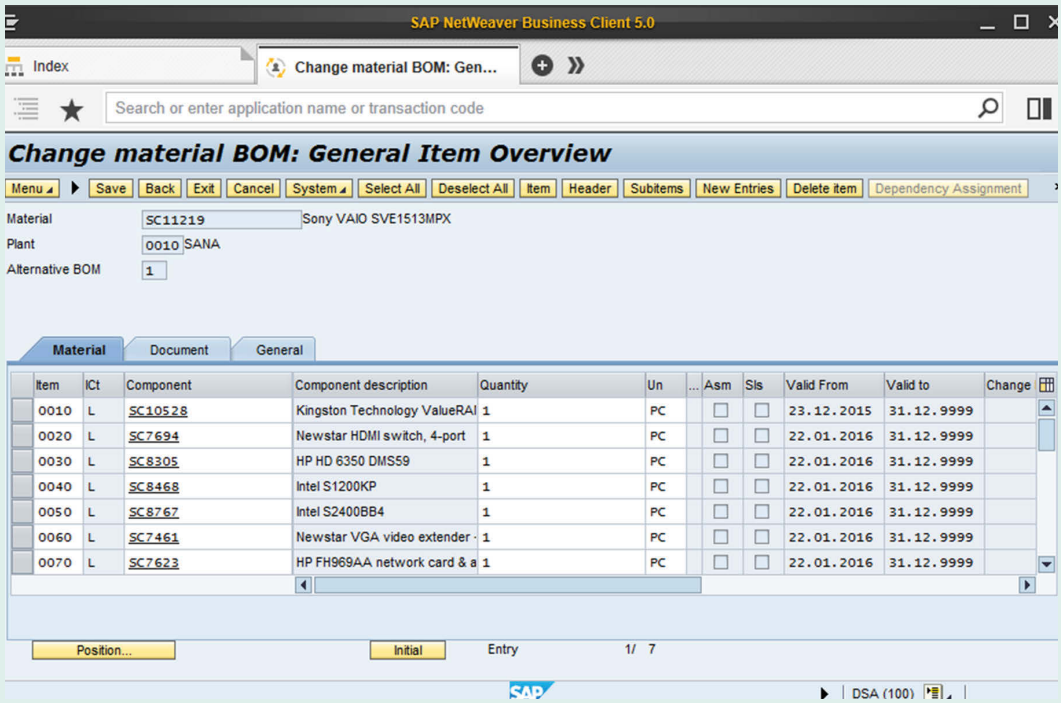


Figure 10.12 SAP interface for changing and updating the "material BOM" (Credit: SAP)

The "material BOM" shown above is only one of many tables that constitute the BOM in ERP systems. For example, in Oracle's ERP products, there are eight "BOM tables," with names that don't require decrypting:

- bom_bill_of_materials
- bom_inventory_components
- bom_reference_designators
- bom_substitute_components
- mtl_item_revisions
- bom_operational_routings
- bom_operation_sequences
- bom_operation_resources

Similar to SAP, Oracle appears to have commingled other data with the bill of materials under the "bom_" prefix.

10.3.5. Working locally with bills of materials

If, for example, your project is to redesign the production line of a subassembly item, you need the information from the production line BOMs in several ways. You would like to do the following:

1. Extract the BOM for this subassembly from the complete product's BOM.
2. Decide how many tiers of the subassembly's BOM to involve in the line redesign.
3. Calculate the required quantities of all the multiple tiers of items going into this subassembly.
4. Flatten the subassembly's BOM, because the redesign of the line may involve changing its own subassembly structure.
5. Introduce a new subassembly structure.

You may find that your ERP system is lacking transactions like "extract subassembly BOM" or "flatten BOM." The best ERP specialists in the company will give you extracts of the current BOM in Excel spreadsheets. Looking for other software tools to work with BOMs on sites like Software Advice, you find tools to edit BOMs, manage revisions, and report on where items are used. Using the additional data attached to a BOM, this software may also help with equipment selection or allocation, or with costing.

Based on their suppliers' claims, if you have the answers to questions like 1 to 5 above, these tools will help you integrate this data with the company's systems but they will not help you find these answers. For this, you are on your own. Let us assume that the ERP team has given you spreadsheets of BOMs, and that you can format them into lists of items and Gozinto statements, you will need to validate the content before attempting anything else.

If the BOM says you need 10 screws to fasten a lid and you really need 20, you can only detect this kind of error by physically assembling a lid. This is a case of the data not matching reality, which you cannot discover from the data alone. On the other hand, you can detect errors that are in the internal logic of the data. Items going into themselves, duplicate Gozinto connections, or loops are identifiable from the data, mostly by simple queries. Long loops, however, are not so easy to detect. With tens of thousands of entries, you cannot do it by examining a graphic visualization of the BOM. You need software to do it for you. You may use, for example, the simple_cycles() function in the NetworkX 2.4 Python package to detect loops in directed graphs.

10.3.6. Making a BOM talk: The Vázsonyi procedure

The most straightforward method for an engineer to extract information from a BOM for such purposes as redesigning a line is to use the Vázsonyi procedure. It entails converting the BOM into a Gozinto matrix and using matrix multiplications and additions. This requires built-in functions from commonly available software packages like Excel, Matlab, or R. "Gozinto" is short for "goes into."

Andrew Vázsonyi was a Hungarian mathematician and operations researcher who also made contributions under the names of Endre Weiszfeld and Zepartzat Gozinto. The Vázsonyi procedure can be used to calculate total requirements on items from the Gozinto matrix. It is clever, simple, and more practically useful with today's IT than it was when Vázsonyi invented it in 1954.[1] Not only were computers

1 Vázsonyi, A. (1954). The use of mathematics in production and inventory control. *Management Science, 1*(1), 70–85.

in their infancy then but the bill of materials concept wouldn't emerge for another two decades. The point here is that an engineer working on designing or improving an assembly shop can use this tool on a laptop to find answers that the ERP system is not designed to provide.

Gozinto tables

In practice, engineers often start formalizing BOMs into Gozinto tables, as they are similar to shopping lists. There are a few differences: Shopping lists don't specify a destination. The entry for "leeks" does not specify that they go into a soup. They include units, as in "2 kg of rice." In BOMs, we keep the units with the item master, not in the Gozinto table, because the unit is the same wherever we use the item. When you generate a shopping list to build a product, you join the Gozinto table with the item master to retrieve the unit. Table 10.2 shows an example of a Gozinto table for product X in Figure 10.9 on page 390.

Table 10.2 A Gozinto table for the product X (a canvas on wooden frame) in Figure 10.9

Item	Goes into item	Quantity	Description	Unit of measure
A	X	1	canvas textile	cm^2
D	B	2	metal stripe	pc
E	B	1	glue	ml
B	X	2	tension clip	pc
E	C	2	glue	ml
F	C	4	wooden list	pc
C	X	1	frame	pc

Converting a BOM to a Gozinto matrix

For small BOMs, there are many tools available to convert a list of Gozinto triplets into a matrix with quantities placed at the intersection of the row for an item and the column for another item it goes into:

- Pivot tables in Excel or Google Sheets,
- Crosstabs in Access or Tableau
- pivot_wider from the R tidyverse package

In most systems, this conversion is one-way. Tidyverse, however, also has a pivot_longer function to convert a matrix to a list of triplets.

First, let's look first at what we can do with a small BOM. Then we can extend it to more realistic, large ones. The following Gozinto graph shows a toy example, in which subassembly S1 is sold as a replacement part for product P1, subassembly S2 is used in both products P1 and P2, and component T3 is used directly in product P1 as well as in subassemblies S1 and S2. Figure 10.13 was generated with the igraph package for R.

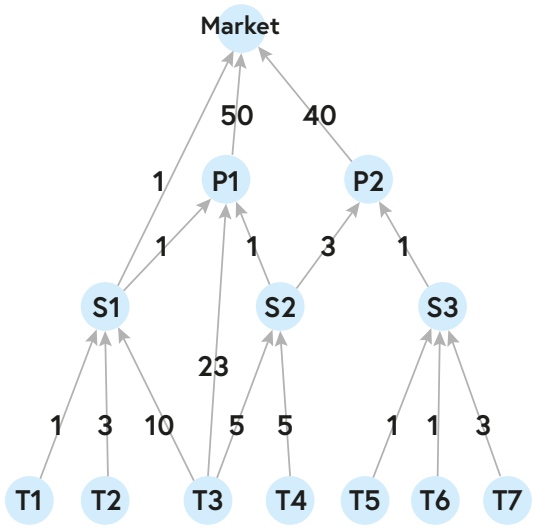


Figure 10.13 An example of a Gozinto graph for a toy

The demand vector is comprised of the items that go into the market. Cross-tabulating the triplets produces the matrix in Table 10.3, with 0s omitted for clarity.

Table 10.3 Toy example BOM as a Gozinto matrix

Item	P1	P2	S1	S2	S3	T1	T2	T3	T4	T5	T6	T7
P1												
P2												
S1	1											
S2	1	3										
S3		1										
T1			1									
T2			3									
T3	23		1	5								
T4				5								
T5					1							
T6					1							
T7					3							

The graph, the list of triplets, and the matrix are three different expressions of the same data. While less visually attractive than a Gozinto graph, the Gozinto matrix enables us to detect characteristics of the BOM, even when it contains thousands of items and we cannot see the whole picture:

1. **Items with empty rows**, like P1 and P2, don't go into any other items and are therefore finished goods.
2. **Items with empty columns**, like T1 through T7 have no other items going into them and are therefore purchased.
3. **Empty diagonal.** The diagonal is, and should always be empty. A nonzero entry in the diagonal

would represent an item going into itself.

4 No direct loops. Wherever there is an entry for row I and column J, there isn't for row J and column I. This means that there are no errors of the type "A goes into B" and "B goes into A."

Calculating requirements with the Gozinto matrix

Let's call Table 10.3, "matrix G." The next step we want to take is multiplying it by a demand vector but the demand we have extracted from the triplets has only three entries, for products P1 and P2 and for subassembly S1 that is sold as a spare part (Table 10.4).

Table 10.4 Toy example demand matrix

Item	Quantity
P1	50
P2	40
S1	1

To multiply this vector by **G**, we need to append a 0-demand for all the items that are not sold on the market, which gives you the demand vector **d**.

When we take the product $\mathbf{G} \times \mathbf{d}$, we get the vector $\mathbf{d}_1$ of the quantities for the level 1 items needed to meet the demand. Then when we take the product $\mathbf{G} \times \mathbf{d}_1$, we get the vector $\mathbf{d}_2$ of the quantities needed for the level 2 items. Finally, $\mathbf{G} \times \mathbf{d}_2 = 0$ because the BOM is only two levels deep. Then $\mathbf{d} + \mathbf{d}_1, + \mathbf{d}_2$ gives you the quantities needed for all items, regardless of level, from finished goods to purchased parts. Note that some items, like the subassembly S1 and the screw T3 appear at multiple levels. Table 10.5 summarizes the total parts needed for the toy example.

Table 10.5 Toy example total parts needed

Item	Demand	Level 1	Level 2	Total
P1	50	0	0	50
P2	40	0	0	40
S1	1	50	0	51
S2	0	170	0	170
S3	0	40	0	40
T1	0	1	50	51
T2	0	3	150	153
T3	0	1,160	1,350	2,510
T4	0	0	850	850
T5	0	0	40	40
T6	0	0	40	40
T7	0	0	120	120

The Vázsonyi procedure

If **I** is the identity matrix, then:

$$\mathbf{d} + \mathbf{d}_1 + \mathbf{d}_2 = (\mathbf{I}+\mathbf{G}+\mathbf{G}^2)\times\mathbf{d}$$

For matrices as for numbers:

$$\mathbf{I}-\mathbf{G}^3 = (\mathbf{I}-\mathbf{G})\times(\mathbf{I}+\mathbf{G}+\mathbf{G}^2)$$

Since $\mathbf{G}^3=\mathbf{0}$, this says that we have a matrix **H** that we can just multiply by the original demand vector *d* to obtain the quantities required at all levels of the BOM:

$$\mathbf{H}=(\mathbf{I}+\mathbf{G}+\mathbf{G}^2)=(\mathbf{I}-\mathbf{G})^{-1}$$

In the toy example, the matrix **H**, with 0s omitted, is shown in Table 10.6. Vázsonyi called the **H** matrix the Total Requirements Factor Table and A.-W. Scheer[2] called it the Total Input Matrix, which we use here.

Table 10.6 The Total Input Matrix (TIM)

Item	P1	P2	S1	S2	S3	T1	T2	T3	T4	T5	T6	T7
P1	1											
P2		1										
S1	1		1									
S2	1	3		1								
S3		1			1							
T1	1		1			1						
T2	3		3				1					
T3	38	15	1	5				1				
T4	5	15		5					1			
T5		1			1					1		
T6		1			1						1	
T7		3			3							1

In general, if *n* is the number of levels in a BOM, then $\mathbf{G}^{n+1} = \mathbf{0}$ and

$$\mathbf{H}=(\mathbf{I}+\mathbf{G}+\ldots+\mathbf{G}^n)=(\mathbf{I}-\mathbf{G})^{-1}$$

2 Scheer, A.-W. (1998). *Business process engineering study edition: Reference models for industrial enterprises* (pp. 120–121). Springer Verlag.

This is the Vázsonyi procedure. If all we care about is the total quantities needed of all items, we can just calculate **H** and multiply d by it. On the other hand, if we are interested in the populations of the different levels, we know that $\mathbf{G}^k \times d$ will give us the items at level k.

Note in particular that a flat BOM – without subassemblies – is characterized by $\mathbf{G}^2 = \mathbf{0}$, and therefore

$$\mathbf{H} = \mathbf{I} + \mathbf{G}$$

10.3.7. Four views of a BOM

We now have four formal ways to represent a BOM that are different ways of organizing, operating on, and visualizing the same data:

1. The Gozinto graph
2. The Gozinto table
3. The Gozinto matrix
4. The Total Input Matrix

In principle, converting between these views is possible. In practice, we often use the Gozinto table as the master, as it is technically easiest to generate all the other views from it. Ideally, we would prefer to create BOMs directly as Gozinto graphs, because a graph with pictures or pictograms of components and subassemblies hand-drawn on a board is a common starting point. It may go down only two levels before getting hopelessly busy but it is intuitive, effective as support for discussions, and unlikely to include structural errors like loops.

Figure 10.14 shows software tools to navigate between these views that an engineer can have running on an ordinary laptop computer:

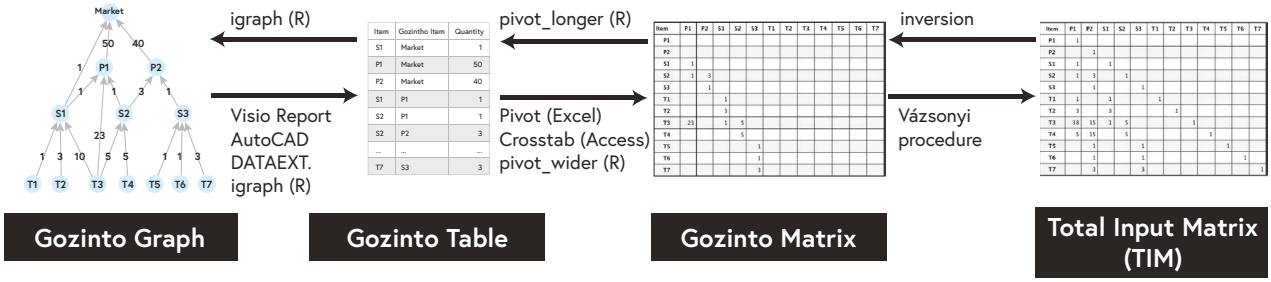


Figure 10.14 Navigating between BOM views

The next step would be to replicate the hand-drawn graph with software, with item IDs and quantities, and automatically extract the Gozinto pairs from this drawing. It can, for example, be done with the following tools:

- Microsoft Visio. If you draw the graph in Visio, you can use a report to extract the Gozinto table.
- AutoCAD has a DATAEXTRACTION command.

The Gozinto graphs in this chapter were generated from Gozinto tables with igraph. The conversion in the other direction, from a graph to a Gozinto table is also possible with igraph.

The main drawback of the Gozinto table is that its manual entry is error-prone. Even for a simple product, missing or extraneous connections become visible only when you generate the Gozinto graph. This happened in the process of writing this chapter: I made a mistake and found out when the Gozinto graph didn't look right.

The Gozinto table is useful for storage, retrieval, and database manipulations. In particular, you can attach differently structured additional data to items in the item master or Gozinto connections between items. For example, a purchased item has a contractually agreed order fulfillment lead time while a Gozinto connection in-house has an assembly process time.

As we have seen, the Gozinto Matrix and the Total Input Matrix (TIM) are more convenient for materials requirements calculations or for reengineering the subassembly structure. Retrieving the Gozinto Matrix from the TIM is technically feasible but usually not necessary. Since, with **I** as the identity matrix, $\mathbf{H=(I-G)^{-1}}$, we have

$$\mathbf{G=I-H^{-1}}$$

Consequently, if ever needed, there is nothing more to retrieving **G** from **H** than a matrix inversion.

10.3.8. Application to LEGO tractors

In 2010, consultant Christophe Caberlon developed the assembly and logistics LEGO tractors simulation game for AGCO's tractor factory in Beauvais, France. This game was based on the 2009 LEGO City Set #7634 that LEGO no longer sells, shown in Figure 10.15.



Figure 10.15 The 2009 LEGO City Set #7634 (Credit: LEGO)

The complete set had 78 components and 47 unique items, including the farmer; the tractor itself, 73 components and 43 items and came in multiple colors, which enriched the simulation game. While LEGO no longer sells this tractor set, you can still buy the bricks it is built from.

The engineering BOM for tractors in multiple colors

The game involves assembling tractors in multiple colors. From the perspective of a design engineer, it makes sense to partition the items between those that are common to all colors and those that are specific to each color. As shown in Figure 10.16, these different sets of items, however, are phantoms, not subassemblies. In the cab, for example, a common part like the windshield is surrounded by colored parts.

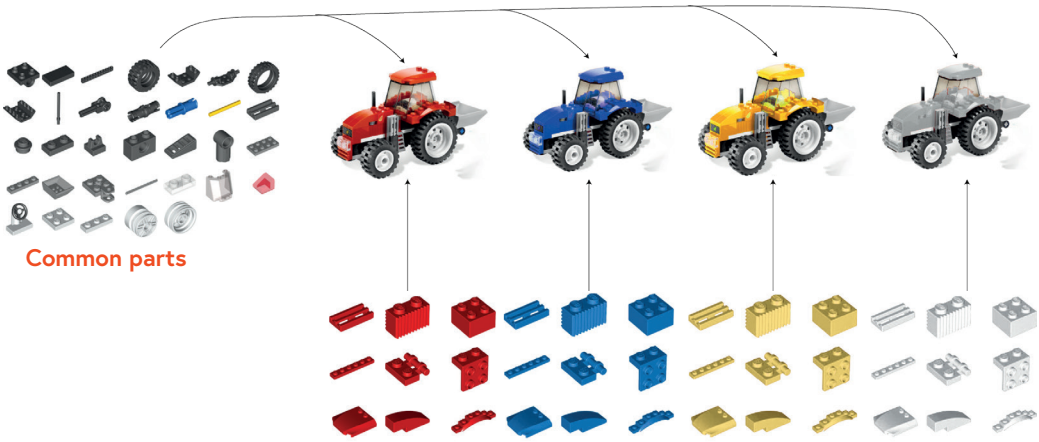


Figure 10.16 BOM with items grouped by color into "phantoms"

Retrieving the LEGO tractor part list

The tractor BOM on the Brickset website[3] is shown as just a list of items and quantities going into the red tractor, with additional data. For human readability, the most useful are photographs and descriptions of the items. The format is shown in Figure 10.17. In a real manufacturing situation, the additional data would include the supplier's lead time and possibly a reference to an agreement setting terms and conditions. For quantity calculations, all we need is the Item ID, Element, and the Quantity used.

To write a BOM with the common and colored parts as phantom subassemblies for the different tractors, we need to set up Item IDs for the blue, yellow, and white tractors as well as for the different sets of parts. As these IDs must be unique, since the LEGO IDs are all numeric, all we need is to start the new IDs with a letter. Because these IDs are not descriptive, we retain the descriptions for human readability:

- A1, A2, and A3 are IDs for the tractors in colors other than red
- A4, for the set of common parts going into all tractors
- A5, A6, A7, and A8, for the colored parts

3 https://brickset.com/inventories/7634-1

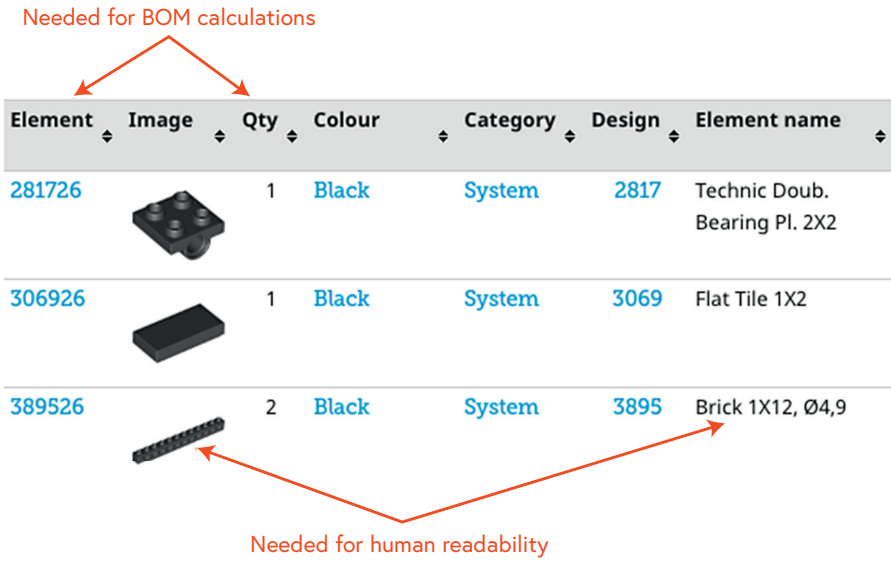


Figure 10.17 Annotated excerpt from the Brickset catalog

The resulting head of the table looks as shown in Table 10.7.

Table 10.7 Gozinto table for the LEGO tractor

Item	Gozinto item	Quantity	Description	Unit
7634-1	Market	32	Red tractor	pcs
A1	Market	16	Blue tractor	pcs
A2	Market	8	Yellow tractor	pcs
A3	Market	8	White tractor	pcs
A4	7634-1	1	Common parts	pcs
A4	A1	1	Common parts	pcs
A4	A2	1	Common parts	pcs
A4	A3	1	Common parts	pcs
A5	7634-1	1	Red parts	pcs
A6	A1	1	Blue parts	pcs
A7	A2	1	Yellow parts	pcs
A8	A3	1	White parts	pcs
281726	A4	1	Technic Doub. Bearing Pl. 2X2	pcs
...	...	...	...	...

To validate the table, we use igraph to draw the Gozinto graph in Figure 10.18. It's not as pretty as the manual one in Figure 10.16, but visually verifying the consistency of the two is a means of checking the table for errors.

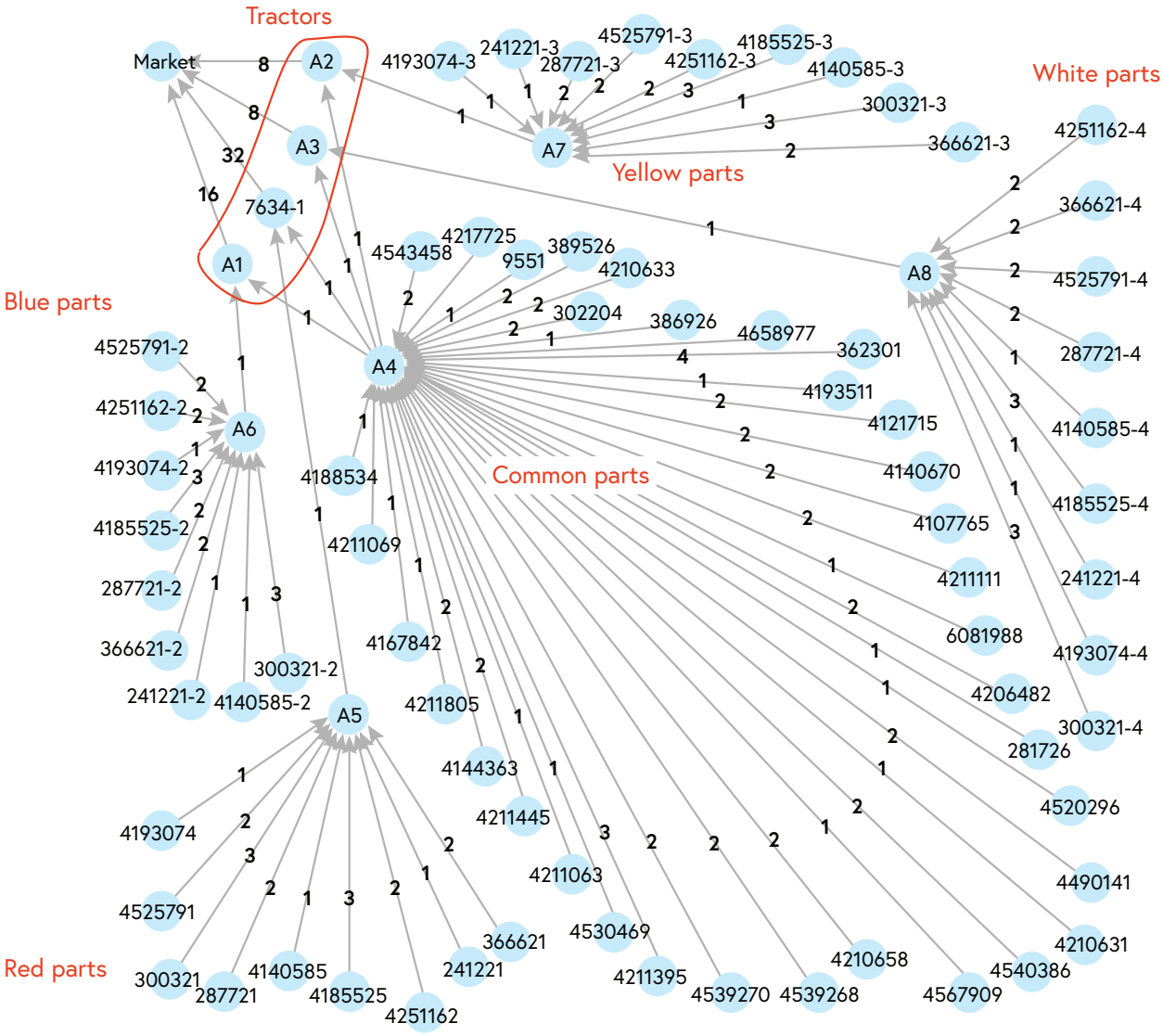


Figure 10.18 Gozinto graph of the Engineering BOM

With 82 directed connections between 80 items, we can barely see the whole structure at a glance and most readers are likely to say: "We're gonna need a bigger sheet." For a real tractor with thousands of items, it is a practical impossibility. The best you can do is draw single-level Gozinto graphs for individual products or subassemblies.

Conversion to a manufacturing BOM

As we have seen, the "subassemblies" in the Engineering BOM are not usable for Manufacturing. First, we flatten the BOM to get rid of these subassemblies. Then we identify subassemblies that make sense for Manufacturing and rebuild the BOM with these new subassemblies.

As discussed above, flattening the BOM takes several simple steps:

1. Convert the Gozinto table into a Gozinto matrix
2. Use the Gozinto matrix to identify all subassemblies as items that go into other items and have other items going into them
3. Remove the rows and columns for the subassemblies from the Total Input Matrix (TIM) to obtain the flat TIM
4. Subtract the identity matrix from the flat TIM to get the flat Gozinto matrix
5. Convert the flat BOM back into a table

Without the phantom subassemblies, the Gozinto graph, as shown in Figure 10.19, looks worse, but we can still see that each common part goes into all the products while each colored part only goes into one product.

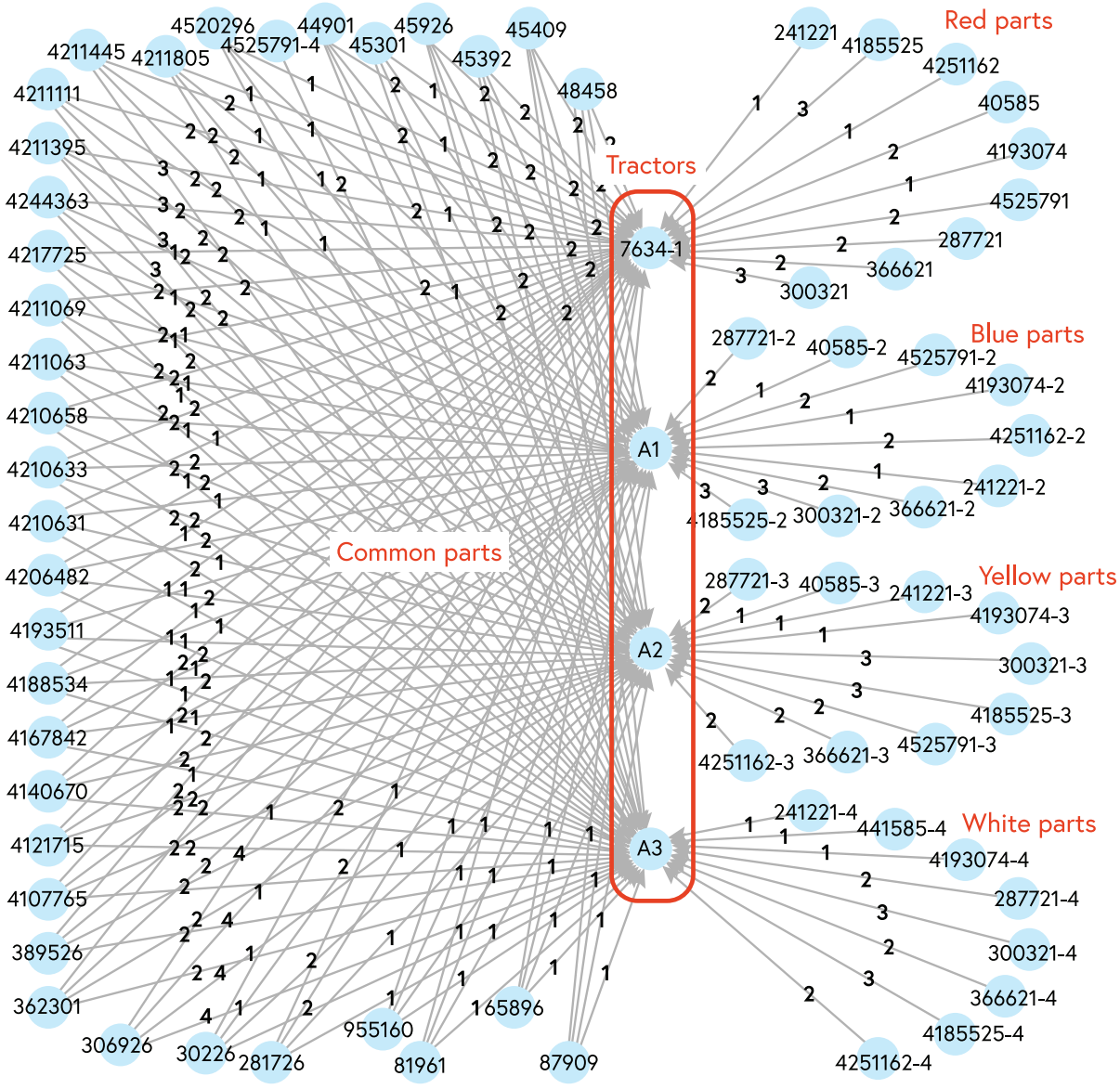


Figure 10.19 Flat BOM Gozinto graph

A common practice to identify potential subassemblies is to fully disassemble and reassemble the product. In Japan, this is called Zenbara (全バラ). If we do it with the red tractor, we get Figure 10.20.

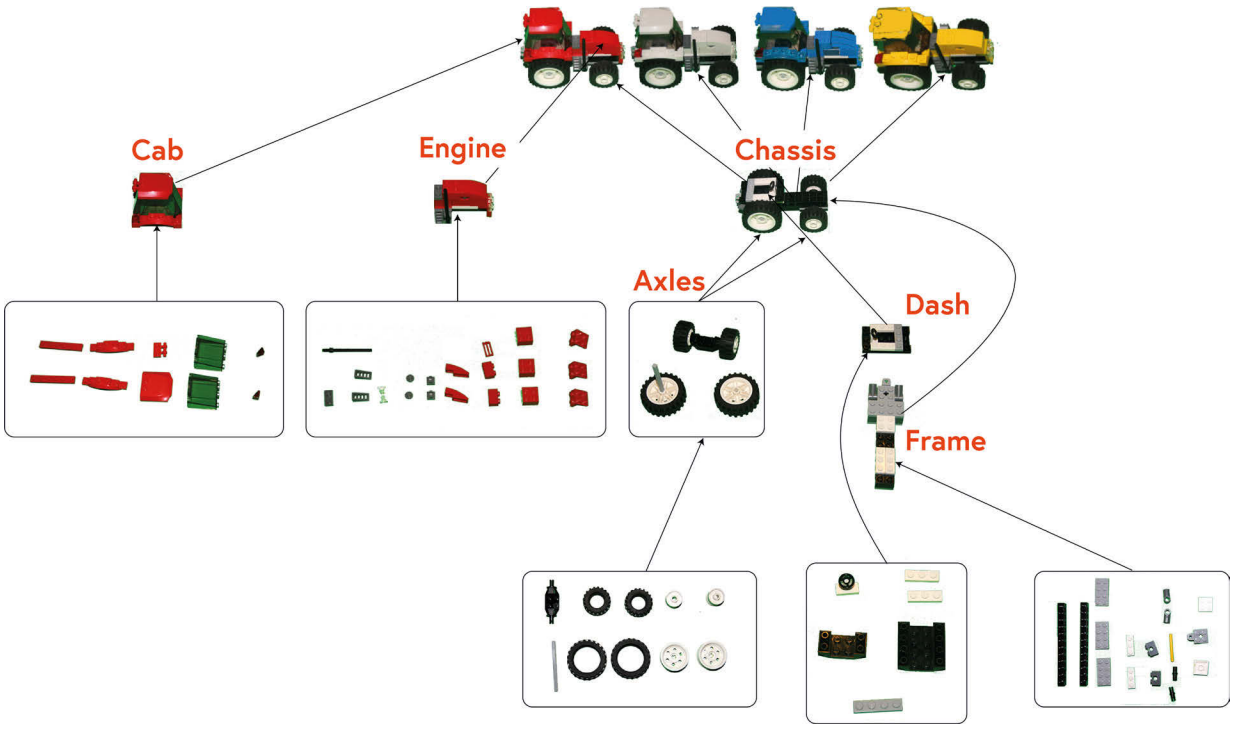


Figure 10.20 Candidate subassemblies

Both the cab and the engine combine colored and common parts; the chassis only uses common parts.

Defining subassemblies that make sense for production is an engineering judgment call involving considerations of tooling and line balancing discussed in Chapter 5. What we are concerned with here is the procedure to integrate a subassembly into the BOM and, for the purpose of describing it, we'll consider the following opportunities:

- Cabs, one of each color
- Axles, front and rear, as common parts
- Wheels, going into the *axles*

After creating the table just for these subassemblies, omitting all the items that still go directly into each tractor, we can draw the Gozinto graph in Figure 10.21. It shows the windshields and the taillights as common parts going into the cabs of all different colors, and the colored parts going into only one. It also shows the front and rear axles going into all the products.

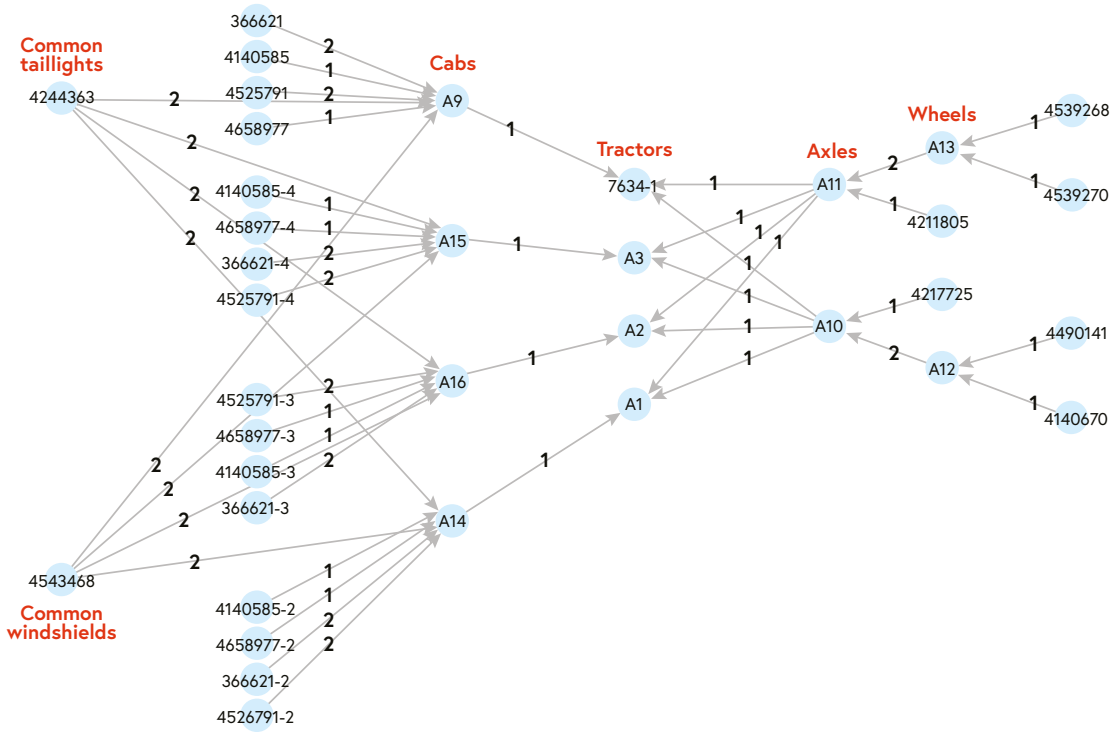


Figure 10.21 Gozinto graph for selected subassemblies

The challenge now is to integrate these subassemblies into the overall BOM. We can always assume we start from a flat BOM and consider only the items that do go into subassemblies, as the other ones are unaffected.

The subassembly Total Input Matrix (TIM) gives us total quantities required of the purchased items to make the subassemblies but each of these items may also be used directly into the finished products. The same screw, for example, may be used both in fastening subassemblies and on multiple finished products.

The steps are therefore as follows:

1. Flatten the original BOM.
2. Compute the subassembly TIM.
3. For each "finished product" and each "purchased items" in the flat BOM, subtract the quantities used in subassemblies from the quantity going into the "finished product" available in the TIM.
4. Filter out the Gozinto records with quantity = 0 from the Gozinto table.

Once we have done this, we have a BOM that can be used to determine how many units of each item need to flow through each assembly operation to produce a given mix of products. Although this is information we need, we must also understand that it is not all we need.

When we attempt to visualize this BOM as a Gozinto graph, we get the busy Figure 10.22. We can still see the subassembly structure, but we can't make out the items that go directly into tractors, even though there are only 30 of them.

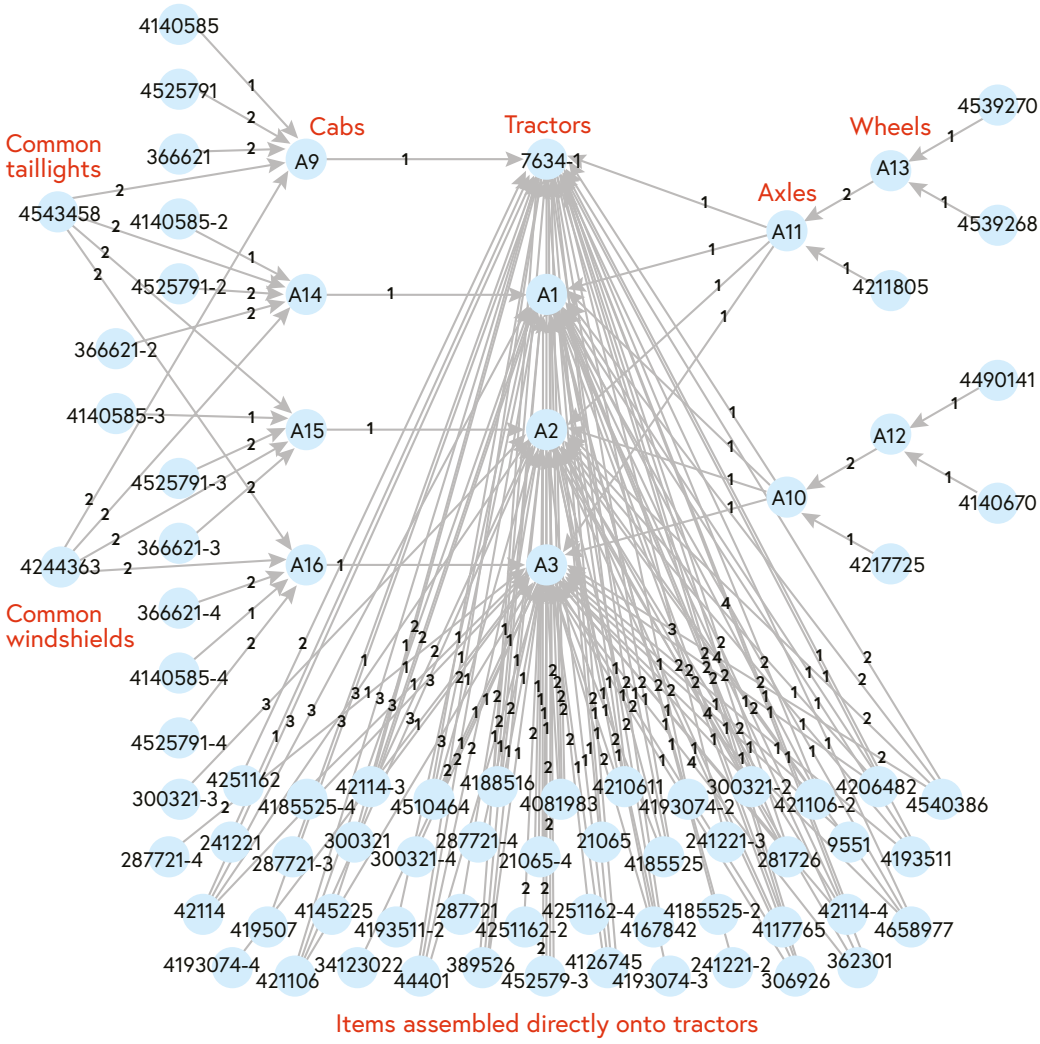


Figure 10.22 Full BOM with subassemblies

Large BOMs as sparse Gozinto matrices

We can explain the Vázsonyi procedure with the toy example above but, in this example, it provides no answers that you couldn't obtain easily from the Gozinto graph. It is useful to the extent you can scale it up to real demands for thousands of products made from tens of thousands of items. Since most items don't go into most other items, the majority of the quantities in the Gozinto matrix are zeros. The BOM of an assembly plant with 20,000 different items would be expressed as a Gozinto matrix of 400 million coefficients, most of them zero. Fortunately, there are ways of representing such a sparse matrix with only the nonzero coefficients and yet operate on them as if all the coefficients were present.

The Gozinto triplets above are in fact a *sparse matrix* representation of the BOM. Software is available to apply matrix algebra on sparse matrices without cross-tabulating them into full matrices but unfortunately not with Excel or any Excel add-ons that we could find. On the other hand, you do find the *scipy.sparse* package to do this in Python or the *SparseM* package in R.

10.4. Equipment master data

Equipment is not all about maintenance. Production machines have process capabilities and capacities, use certain tools, fixtures, and jigs. They also have startup, shutdown, and setup procedures, and they require qualified operators. All this needs to be taken into account and master data about equipment must be maintained just as much as master data about the products and processes.

In addition to production machines, a factory's equipment includes many types that are described by different data. If we take just the example of their locations, we have as the following:

- **Plantwide systems**, like air handling, electric power supply, compressed air, fluids provided through pipes, conveyor networks, or drainage systems for waste. For production, these systems are described by their interface points, like inlets and outlets; for facilities and maintenance, by maps showing all their nodes and connections.
- **Cranes** have a range and weight limits.
- **Machines** have a location that may change as line layouts are revised but not during normal operations. They may have technical requirements that restrict users' ability to relocate them, for example a 20-foot concrete foundation for stability.
- **Dies, fixtures, and jigs** have permanent storage locations, from which they move to specific machines as needed.
- **Vehicles**, like forklifts or tugger trains, have a parking space and authorized routes.

In terms of master data, each item of equipment needs the following:

1. A unique ID and a list of characteristics of interest based on the uses it is put to. As for products, the only job of the ID is to be unique and no information needs to be embedded in it.
2. For production, links to the operations the equipment is capable of performing, including tooling, programs, instructions, and qualified operators.
3. For maintenance, links to first responders to problems, skilled technicians, and outside support services, from the supplier or distributor.
4. For communications and control, the protocols used to exchange data with the equipment.

10.4.1. Generic and specific equipment states

From the time a plant orders a piece of equipment to the time it disposes of it, it undergoes many state transitions. Some of these states are generic and applicable to all equipment. For example, it may go from "on order" to "received," and then "installed," "in-testing," 'in production," "on standby," etc. Within "in production," it has sub-states like "running," "idle," or "down."

The generic states and the events that cause transitions between them are part of the master data and provide a common language for management to view the overall state of the equipment at one time or the performance of one piece of equipment over time.

Tower lights are often used to make the state of each machine visible but the meaning of the different light colors is inconsistent between machines and there are often too many colors and color combinations for humans to remember. A generic state model is necessary for consistency so that, everywhere you look, red means down, yellow idle, and green working. You make sure that every tower light has

one and only one light solidly turned on at all times and that operators do not have to learn the meaning of, for example a solid white with a blinking blue.

Computerized Maintenance Management Systems (CMMS) usually come with a built-in model of generic equipment states. In IBM's Maximo system, for example, a piece of equipment is called an "asset" and it is either "not ready," "broken," "decommissioned," "inactive," "missing," "sealed," "active," "limited use," or "operating." It is easier for a CMMS supplier to provide a one-size-fits-all model of equipment states than the capability for each plant to define its own, but, in any given plant, this standard model is likely to be too detailed in some areas and not enough in others. The needs are not going to be the same in a chemical plant and a vacuum-cleaner assembly plant. Plant managers should analyze them internally prior to evaluating CMMS options and possibly favor a system that lets them define their own equipment states and transitions.

In addition to generic states relevant to all identified pieces of equipment, there usually are more specific states like "pressure warning light on" that are applicable only to some classes or pieces of equipment. This level of detail is part of the master data of the equipment controller but perhaps not of the plant as a whole.



Figure 10.23 Equipment master data must be maintained just as much as material master data

10.5. Routings master data

With routings, the focus switches to the physical transformations of materials and components into products. As for production BOMs, routings are assigned to the end product. A routing holds master data that captures the process requirements of a given item. The master data should be structured to cleanly separate the transformations to be achieved from the means of achieving them.

10.5.1. Process breakdown and validation

As discussed in Chapter 4, at the highest level of abstraction, a routing breaks down the transformation into a sequence of states that workpieces undergo on the way from raw material to finished products. Ideally, engineers would go through the following sequence for each transition:

1. Determine how it can be validated, using gauges, sensors, or instruments.
2. Choose a physico-chemical process to effect it.
3. If assembly is involved, include the relevant BOM records.
4. Design or select the tooling: Utilities, fixtures, jigs, and consumables.
5. Code the process program.
6. Design the operator tasks required.
7. Select, buy, or build the equipment.

The ability to validate the transformation of a workpiece is a key criterion in breaking down a process but it is not always possible. In machining you can check critical dimensions or weight, or run leak tests. In assembly, you can check for missing components or fastening strength. In semiconductor wafer processing, on the other hand, in-process checks are of limited value because the critical characteristics set by an individual operation are often not observable at the completion of the operation. The entire process ranging from 200 to 600 process steps is only validated at final test.

There is often more than one possible process, different tooling options, multiple ways to design the process programs, and multiple types of machines capable of executing it. It is not uncommon to find, in the same shop, machines of different vintages with overlapping capabilities and different capacities, flexibility, setup structures, etc.

The breakdown of a process is often hierarchical, with each level given a different name. A machining process may have phases like roughing and finishing, or soft-machining, heat treat, and hard-machining. Each phase may then break down into a sequence of operations, which themselves breakdown into steps. The number of levels and their names may vary between organizations. In this tree, the higher levels are generic, with the same sequence applying to many products. The lower levels are increasingly specific to each product.

Figure 10.24 shows an example of how a generic routing looks like for a process in SAP software. The first column shows the operation number, the column labeled “Description” shows the high level process, and the column “Machine” indicates the default machine dedicated to the process.

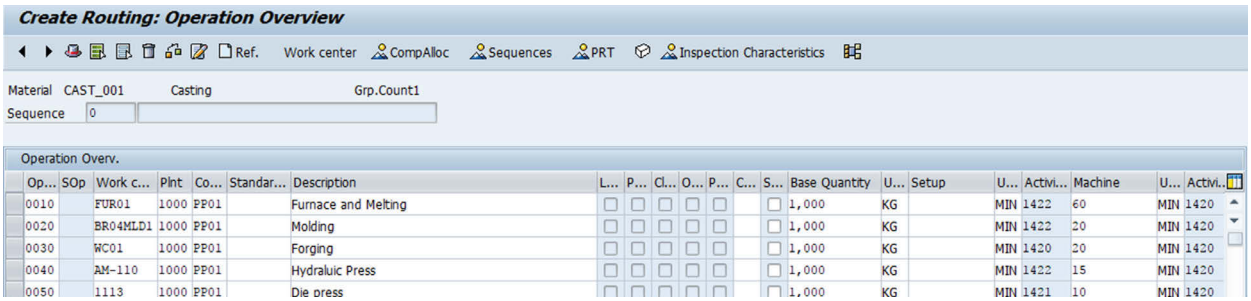


Figure 10.24 Example of creating routing master data in SAP

10.5.2. The edge

The edge is where the routings meet production operators, tools, and machines and where all the lowest-level details come together. In production lines, the data side is in the form of instructions at each station and process programs downloaded to machines.

The operator instructions cover part picking or kitting, assembly and the manual work done on machines, and validation of the work through go/no-go gauges or measurements. If kitting is fully automated with robots or dispensers, or partially automated with pick-to-light systems, the pick lists are downloaded to the relevant control system.

Making all these elements converge to each workstation for each workpiece is a technical challenge, but meeting it is not sufficient. When operators repeat the same routine hundreds of times per day for months, they "go on automatic," meaning that they go through the motions with their minds disengaged, which means that any change requires them to unlearn some habits and learn new ones. Unless the system, in cooperation with the leaders, brings the change to their attention, they will keep executing the old methods.

10.5.3. Learning from paper travelers

In a line, the sequence of operations is materialized in the physical layout. In a job-shop, the physical location of a workpiece does not tell you where it is in its process, and this information has traditionally been provided in a document called "traveler" that stays with the workpieces throughout the process, is used to collect data, and is retained for traceability.

While obsolete in the digital age, paper travelers are still found in many factories and are a good source of requirements for the routing information model. They reveal, for example, that operations do not neatly fit in a table because the parameters describing each operation vary in nature, number, and complexity. They also reveal the level of detail to which each process is documented.

10.6. Locations master data

The layout and maps of a manufacturing site is part of its master data, and the locations of all activities are essential data to support operations. The buildings on the site usually have names that are no help in finding them. Some are named in honor of distinguished past employees; often they are numbered in order of construction. The South San Francisco complex of Genentech has building numbers going up to 85, but buildings with consecutive numbers are not necessarily adjacent or even close. Buildings 11 through 13 are contiguous but buildings 35 and 36 are not. This does not help people or goods move within the site. The issue is that a site is *two-dimensional* and a numbered sequence *one-dimensional*.

Modern manufacturing buildings are usually single story structures with the roof supported by a grid of columns, and grid coordinates painted high on the column, making it easy to find a production line between columns H5 and H8. In older factories, some locations have historical names like "the railroad platform" because, 30 years before, this location had been used to receive materials from trains. Today, such legacy names are confusing.

The master data must include but not be limited to layouts and maps in graphic form. It must also include the grid locations of production lines, warehouses, aisles, docks, facilities systems, etc. as directly accessible records that can be used to route materials or people.

10.6.1. Location of storage areas

While the grid of building columns directs a vehicle to a warehouse, it is not sufficient to locate storage locations within the warehouse. The exact addresses within each warehouse are master data often maintained in Warehouse Management Systems (WMS), which are usually more popular with their users than any of the other computer systems used in factories.

Decisions about the sizing, location, and internal organization of warehouses, however, affect the dock-to-dock flows of materials and have other stakeholders than just warehouse managers. For example, storage strategies like block stacking, dedicated gravity flow racks, single-deep pallet racks, or supermarkets on the shop floor, must be reviewed and approved at the level of operations management.

Technically, production can treat materials management/internal logistics like a supplier, who responds as it sees fit to requests for materials. Just because it can be done this way, however, doesn't mean it should be. In particular, it would imply that the warehouse master data would be private to warehouse management.

In single-deep pallet racks, slot addresses include aisle numbers, column numbers, and height level. Other forms of warehouse organization may need more or fewer coordinates. These addresses, however, are best decoupled from physical locations. Slot 1-5-3 may currently be a column grid location H5. If management relocates the warehouse, the same slot may end up at location B12 but it can still be known as slot 1-5-3, and storage or retrieval directions should include both the physical location and the slot addresses.

10.7. External relations master data

Manufacturers maintain relationships with external entities such as suppliers, contractors, customers, and alliance partners. To keep accurate records of names, addresses, and contracts is of course an essential activity for any firm. If the invoice does not make it to your customers, you will not stay long in business. This data is maintained in the ERP system.

The master data on external relationships includes policies and agreement templates that purchasers, buyers, and alliance managers are expected to use, and all the agreements actually made. As each external entity has its own templates, the actual agreements are the result of negotiations and often include accommodations that deviate from both sides' templates.

For contracts, only a small fraction of the data is used daily. It includes procedures for issuing and accepting orders, shipping and receiving goods, etc. It should be structured and kept readily available. I should also include the procedure for handling the most common exceptions, like a truck delayed in traffic. Most of the contracts consists of data that both parties hope will never be needed, and is intended to cover the handling of all possible exceptions and the resolution of all conflicts. The full contracts are text documents kept in document databases or data lakes, from which most are never retrieved. They need to be kept just in case they are needed, but they don't need to be at every clerk's fingertips.

10.8. Relationships between master data

The classes of master data described above, and other master data, have mutual relationships that must be maintained as well as their individual values. For example, as discussed above, Routings are sequences of Operations but the same Operation may be used in more than one Routing, and Operations are individually engineered prior to being integrated into Routings.

In principle, if an Operation is capable in terms of its critical quality characteristics, it should work as part of a Routing. In reality, the list of characteristics is never complete, and changes in one Operation may have unintended consequences at Operations in a Routing. A change in a milling Operation on forgings may generate new stresses in a workpiece and causes it to warp at later Operations. For this reason, any engineering change to any Operation includes integration testing in all the Routings that use it.

To support overall process engineering, the master data must include the many-to-many relationships between Operations and Routings. Database technology can help maintain the integrity of this data, for example by disallowing the inclusion of an undefined Operation in a Routing but it is much less adept at preventing users from defining the same Operation multiple times with different names as in the example in Box 10.5.

Box 10.5 **Intel's copy exactly strategy**

Intel, for example, had the problem of nomenclature inconsistencies into the 1990s. As it deployed the same processes to make the same products in multiple factories worldwide, the same operation was called "Oxidation 1," "Oxide1," and "Oxide 1" in different factories. With Routings of 500 Operations, engineers could neither cross-reference technical performance data on this operation from multiple plants nor coordinate the deployment of updates.

The technology of the time made it possible to maintain all the master data of a process in a shared database. Intel's management provided the determination to use this capability in the form of a program called "Copy Exactly!", now known as CE!, in which a process, developed in a lab, is rolled out to production in multiple facilities with identical equipment sets and a shared database, not only for master data but also for status and historical data.

The most commonly used systems in Manufacturing keep their data in relational databases. That is in sets of tables where relationships are established by matching values in columns. In the above example, the Routings table has a column called "Operation" containing the IDs of rows in the Operations table.

By explicitly designating a relationship as a link between tables, you can ensure that no Routing contains a non-existent Operation, that changes to an Operation ID are automatically propagated to the Routings that use it, and that deletion of an operation causes the deletion of the records that contain it in Routings.

The more recent "NoSQL" database systems have been criticized for not having a similar way to express relationships, with the exception of graph databases, which elevate relationships to the status of objects on an equal footing with the entities they relate. With Operations and Routings, in a graph database, you would have a third object that you might call, for example, Operations-to-Routings-map to tell you which Operations are used in a given Routing and which Routings share an Operation.

When discussing the relationships that need to be maintained among master data, a graphic representation helps. There are many graphic conventions available, known as Entity-Relationship (E-R) diagrams. They take some learning to use and understand but let you visualize at multiple levels of detail the relationships that need to be maintained among the different categories of master data, in other words, the information model of the plant.

10.9. Master data revision management

A factory's master data is, by definition, valid until management decides to change it. These changes are not usually made by an individual acting alone. The change procedure is usually similar to the following:

1. An individual or a team initiates a change by submitting a request for approval.
2. A group of stakeholders reviews the change request.
3. By approving it, the reviewers trigger several events:
    a. Technical Data Management incorporates it in the master data.
    b. Technical Data Management archives the previous version.
    c. Production implements it on the shop floor.
4. By rejecting it, the reviewers make the initiators either revise and resubmit it, or abandon it.

This short description shows the essence of the process but glosses over many difficulties. The challenge is to make changes possible while providing enough checks to protect production against unintended consequences.

Revision management often face the following issues:

- **"Temporary" workarounds.** Changes are often made in emergency situations, with no time to follow the procedure. In these cases, they are implemented first, as "temporary workarounds" and reviewed later, with the obvious risks or unintended consequences and of the proliferation of workarounds causing the actual process to stray.
- **Congestion.** It is not uncommon for a factory to have hundreds of pending engineering changes and for the approval process to take months. The actions required to relieve this congestion are beyond master data management. They have to do with product and process design, and involve many stakeholders with objectives and concerns that are not easy to reconcile. What master data management can do is not add any more friction to this process.
- **Multiple approval stages.** Engineering change procedures often have more than one stage of approval. The first stage, for example, may give permission for the engineers to proceed with experimentation. The intent is to quickly filter out half-baked proposals but the result may be additional delays.
- **Timing changes.** Most changes are made effective as of a given time. When they are technical changes to a manufacturing process, however, effectivity by time may not be appropriate. If, for example, one change affects two operations within a route, making it effective as of a given date may cause some workpieces to have the first operation done according to the old version and the second operation according to the new one. Generally, decisions about effectivity management should be driven by such technical considerations, not by the level of information technology used to execute it. If, however, the plant uses paper travelers, then effectivity by time would require updating all the paper travelers in progress. In such cases, it is easier to manage effectivity by serial number than by time, even when the process does not require it.
- **Storage and retrieval of obsolete versions.** For traceability and analysis purposes, the obsolete version after a change should be archived and linked to all the product units made with that version.

Summary

The most essential points in this chapter were:

1. Master data management is an essential but difficult job.
2. Master data are timeless and include both lists of items and relationships.
3. Today, master data management is managed using IT.
4. Nomenclatures like "smart" product identifiers are not as smart as they sound. The key approach is a better alternative.
5. The bill of materials (BOM) is a comprehensive list of parts and the quantities needed to manufacture a complete product and its subcomponents, including their relations.
6. A Gozinto ("goes into") table lists all parts and subparts and the parts they go into.
7. Equipment, routing, and location master data must be managed.
8. Master data about suppliers, customer, and alliance partners must be managed.
9. Relationships between master data must be defined and maintained.
10. Master data revision management follows a rigid approval process.

Test yourself

1. What separates master data from other data?
2. What is nomenclature and how does smart data differ from the key approach?
3. What is a BOM?
4. Why are BOM data often inaccurate? How can we measure BOM accuracy?
5. What are some typical parameters stored as master data of equipment?
6. How does routing master data differ in job-shop versus a production line?
7. What is meant by "the edge" in routings?
8. What external data would you store as master data?
9. How can relationships between master data be identified?
10. What are some of the main reasons why master data revisions are inefficient?

Further reading

Allen, M. & Cervo, D. (2015). *Multi-domain master data management: Advanced MDM and data governance in practice*. Morgan Kaufmann.

Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., van Run, P., & Wolfson, D. (2008). Enterprise master data management: An SOA approach to managing core information. Pearson Education.

Kloock, J. (1969). *Betriebswirtschaftliche Input-Output-Modelle*. Gabler.

Loshin, D. (2010). *Master data management*. Morgan Kaufmann.

Mather, H. F. (1986). Design, bills of materials, and forecasting – The inseparable threesome. *Production and inventory management, 27*(1), 90–107.

McCracken, V. (2016). *Requirements for an MDM Solution: A proven approach for how to gather, document, and manage requirements for a Master Data Management solution from inception through implementation*. CreateSpace Independent Publishing Platform.

Treder, M. (2020). Masterdata management. In: *The Chief Data Officer management handbook*. Apress.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Part 4

Making Materials Flow

DOI:10.4324/9781351110310-14



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

11 Internal Logistics

In manufacturing, logistics encompasses all material-related operations except actually transforming materials. *Internal* logistics covers those that occur *within* the production facility – that is, the dock-to-dock material flows inside the plant, from receiving to shipping, including part presentation to assembly. In internal logistics all the work is done under the authority of a single business entity. In contrast to external logistics and supply chain management, where independent business organizations coordinate or collaborate.

In internal logistics, materials move in quantities that are orders of magnitude smaller than in the supply chain but at frequencies that are orders of magnitude higher. Internal logistics is to production what the pit crew is to a race car driver. Its first objective is to deliver the right materials to the right locations, in the right quantities, at the right time, and in the right presentation; its second, to do all of it efficiently.

Parts usually spend weeks in the custody of a logistics organization for only a few hours in production. As a consequence, one may ask why so much effort should be directed to improving production when there is more at stake in logistics. This question, however, is predicated on the mistaken assumption that the two are unrelated. Logistics performance is largely determined by what happens in production, even if logistics accounts for most of the manufacturing lead time.

This chapter discusses how to achieve flow in internal logistics, including both the physical organization of transportation – conveyor systems, small trains, automatic guided vehicles (AGVs), gravity flow racks, forklifts, etc. – and the logic of the protocols used to trigger actions and monitor the material flows.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Define internal logistics.
2. Understand the importance of flow in internal logistics.
3. Use the concept of push versus pull production.
4. Design solutions for pull signals.
5. Decide when to kit.
6. Select solutions for part transportation.
7. Select solutions for part presentation.
8. Choose the right logistics technology for factory processes.

DOI:10.4324/9781351110310-15

11.1. What is the scope of internal logistics?

By definition, logistics does not transform materials. Some people erroneously conclude from this that there is no value added in logistics. Logistics authors counter that logistics provides the value of time and place. In addition, we also see value in *presentation*. Logistics adds three types of value:

- **Time.** A product is more valuable when you need it than a week later.
- **Place.** A product is more valuable where you are than far away in an unknown location.
- **Presentation.** A product presented to you for easy identification and access is more valuable than one mixed in a box with many other parts.

Production does not only run machines and assembly stations, but is also responsible for the flow of workpieces between contiguous stations. Even though these transfers do not transform the workpieces, this activity is still considered to be production, because it is run by the Production department. Conversely, palletizing goods in shipping might be construed as transforming them, but it is still considered part of logistics because it is done by an organization called "Logistics," "Materials handling," or "Materials management."

Where the boundary is placed between Logistics and Production is a managerial decision. Technically, part preparation by Logistics overlaps with the materials handling done by Production, and the handover can take place at many points. The extent and boundaries of the logistics space in manufacturing are shown in Figure 11.1. While logistics activities take up much more space than production, they usually employ an order of magnitude fewer people.

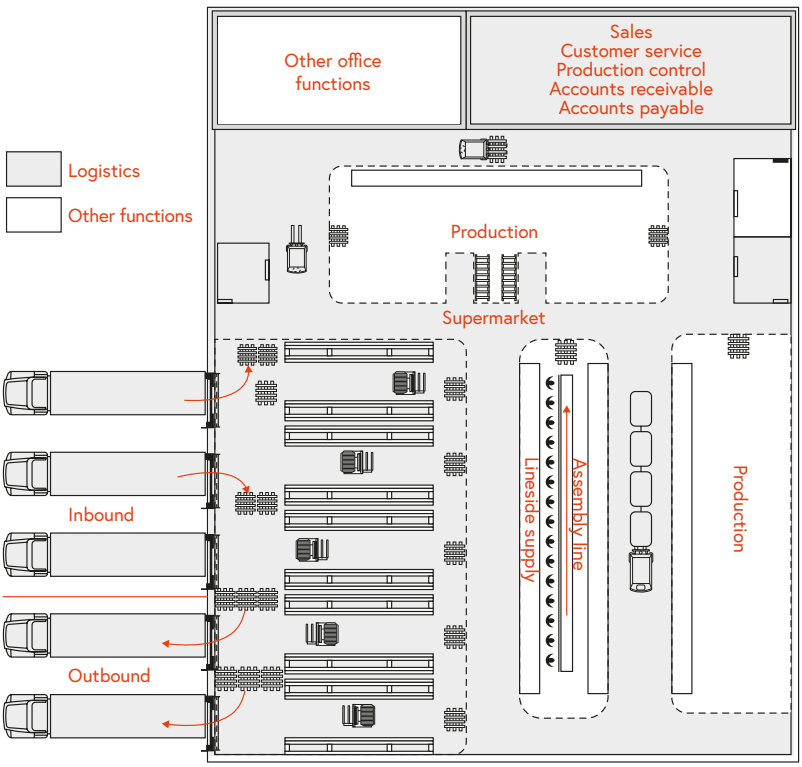


Figure 11.1 Scope of internal logistics

At some locations, materials handlers deliver open bins of parts to flow racks directly behind the assembly line; at other locations, to a supermarket from which a special operator called "water spider" working for Production prepares the parts, picks kits, and delivers them to assembly. In the first case, the boundary between Logistics and Production is right behind the assembly station; in the second case, one step removed from it.

Another key boundary is that between the plant and the rest of the world, which is usually materialized in docks for receiving and shipping. Hence, *internal logistics* is often also called *in-plant logistics* or *dock-to-dock logistics*. Besides the obvious differences in distances, quantities, and vehicles, internal logistics differs from the *in-bound logistics* of getting parts from suppliers and the *outbound logistics* ("shipping") of distributing finished goods in the way it is managed. Internal logistics is under the control of one organization.

External logistics, on the other hand, is ruled by the interaction of multiple independent economic agents making their own decisions, including multiple tiers of suppliers and distributors, trucking companies, railroads, and air and sea freight companies.

11.1.1. The one-size-fits-all approach is illusive

Many manufacturing organizations use only one way to do the work of logistics. One common pattern for inbound operations inside the plant is as follows:

- Parts arrive in full truckloads from one supplier in each truck.
- Operators use forklifts to unload pallets from the truck.
- After the receiving operations, forklift operators put away pallets into single-deep pallet racks into any available slot, and log the location with a radio terminal.
- The Manufacturing Execution System (MES) or the Warehouse Management System (WMS) issues workorders or routing slips.
- The forklift operators retrieve full pallets from the warehouse and log this operation with a radio terminal.
- They deliver full pallets wherever they can find space to set them down near the destination production area.

The same pattern is applied to *all* items, regardless of required quantities or frequency of use. The logistics organization is attached to this one-size fits-all approach, and its management is particularly concerned that its members would not be able to deal with the complexity associated with a different approach for each category of items.

This approach fails to take advantage of varying performance requirements. The key message about in-plant transportation is that the system needs to be customized to the needs of the plant, using a variety of methods. Using only one, as many plants do, is the easiest to set up, but it results in using the vehicles that are suited for the most demanding task for every task, which is wasteful and ineffective. There is no one size that fits all in logistics.

11.1.2. In-plant transportation

The identification of the most heavily traveled routes – and measurements along these routes of parameters such as the number of times the materials are touched or the number of times they move vertically and horizontally – reveals improvement opportunities that are obvious in hindsight:

- Besides improving safety, restricting forklifts to marked aisles may be the key to massively increase the manufacturing density of the shop floor.
- Relocating a frequently used item may be enough to change its handling method from a forklift with a certified driver to a pallet jack hundred times cheaper that anyone can use.
- Changing the minimum in-plant transportation quantity from a pallet to a small bin may eliminate the returns of partial pallets to the warehouse for incoming materials, while simplifying and accelerating the movement of finished goods.

The analysis may also reveal that decentralized manual palletizing with semi-automation retrofits may be cheaper and more reliable than one central automatic palletizer fed through a network of conveyors. The solutions frequently include a transition from forklifts dispatched by radio, like taxis, to pushcarts or tugger trains operating on fixed routes at a regular pitch, like buses and subways.

Regardless of distance, the following needs to happen:

1. The minimum transportation quantity must accumulate at the point of origin.
2. The parts need to be prepared for transportation, which may entail, for example, placing them in bins and palletizing the bins.
3. The part must be transported, for example by a person, a forklift with a driver, or conveyor.
4. At the destination, most parts must be prepared for production – that is, removed from pallets and bins and possibly placed on lineside shelves or kitted.

Cutting the distance between two lines in half will make a difference, but integrating the two lines and eliminating the transportation step will make a much larger difference.

> **Stop and Think 11.1**
> Use the everyday analogy of trains, buses, taxis, and own car/bike to think about how internal logistics for groups of items in a factory may be differentiated.

11.1.3. Transportation aisles

Many shop floors have no marked transportation aisles. This has three main consequences:

- Extra materials, pallets, fixtures, surplus equipment, and other items encroach upon and eventually block transportation aisles.
- There are no restrictions on where vehicles are allowed to go, and production operators must keep looking over their shoulders to make sure no forklift is coming.
- Any visitor who is not familiar with the usual truck driving pattern is in high risk.

Figure 11.2 shows how the shop floor needs to be organized into manufacturing islands – where only production personnel work and into which vehicles from the materials or logistics group may not enter

– and transportation aisles that are reserved for vehicles; to cross operators must exercise the same caution as pedestrians on a public roadway.

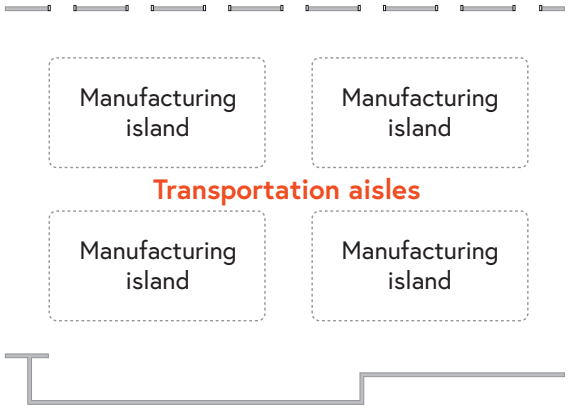


Figure 11.2 Network of transportation aisles

Aisle markings should be kept current and standardized. The adherence to them should be enforced. This is easier said than done when the floors are trafficked by heavy forklifts hundreds of times every day. A few markings that are inaccurate, obsolete, or ignored are sufficient to destroy the credibility of all markings with production operators and transporters alike.

Figure 11.3 shows examples of acceptable aisle marking. If enforced, the crosswalk on the left side is an improvement over random crossings of the transportation aisles by operators. Layouts that avoid aisle-level crossings altogether are safer and more effective but a design challenge.



Figure 11.3 Marked transportation aisles and no-go zones for trucks

11.1.4. In-plant milk runs

The need to move small quantities of a large number of items within plants with short, predictable lead times and without multiplying transportation costs has driven manufacturers to implement so-called milk runs. Milk runs are loops through a number of pickup and delivery points. It is a reference to the services that delivered fresh milk to homes in the 1950s. The term was first used in the *Electric Railway*

Journal in 1917 for a train that collected milk from farmers along a 40-mile route from Morenci, MI to Toledo, OH. The term later became aircraft pilot slang for an easy flight, like a "cake walk."

In manufacturing, the concept is more known as a form of inbound logistics from a network of supplier plants. Manufacturers who organize their own incoming supply can setup pickups and deliveries at fixed times along fixed routes. This use of milk runs originated in the German military aircraft industry in World War II. The point was to avoid the accumulation of stocks vulnerable to enemy air raids at aircraft assembly plant. The Japanese term for this approach is *junkai* (巡回), meaning "tour" or "patrol."

Other metaphors could have been used to describe this approach. Inside a plant, dispatching forklifts on request to pickup locations can be viewed as a taxi system, in opposition to which milk runs can be viewed as buses, picking up and dropping off passengers at a series of stops at fixed intervals on regular routes.

As shown in Figure 11.4, the milk run concept applies in different forms to inbound, outbound, and in-plant logistics, at least for some of the items consumed or produced.

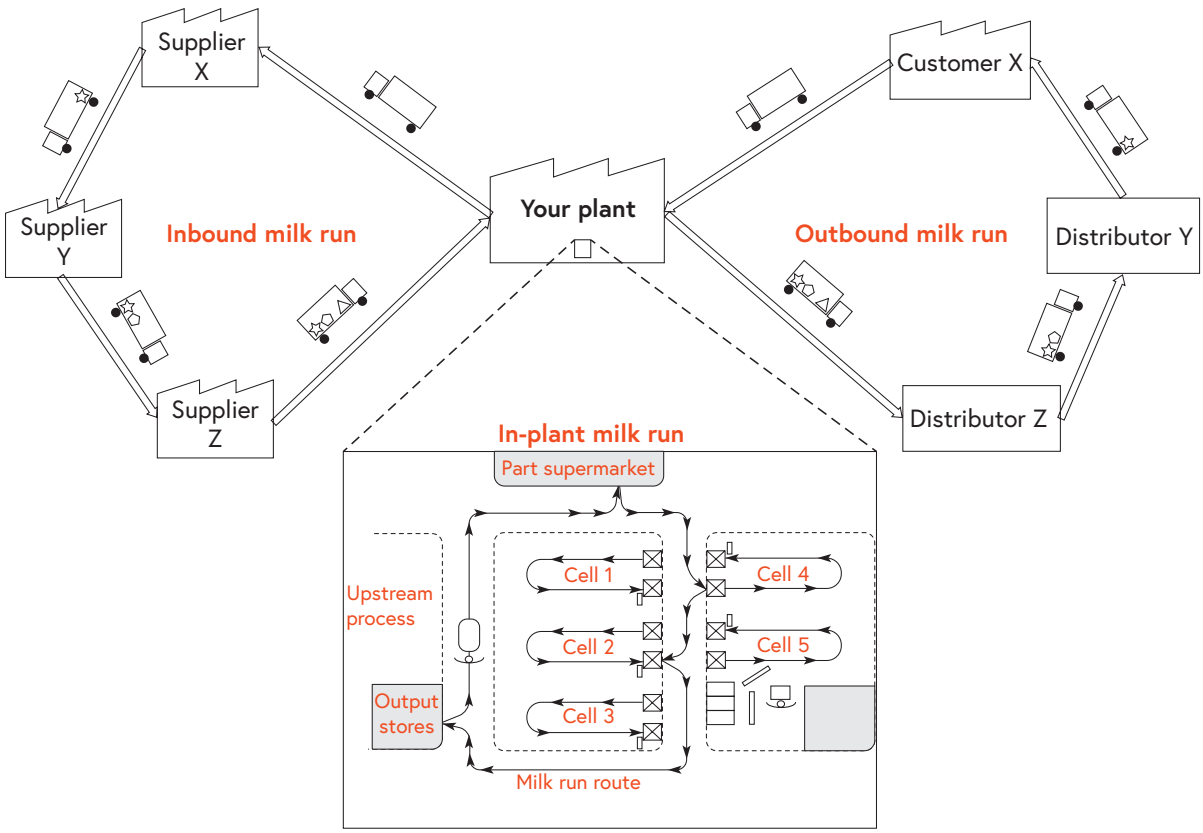


Figure 11.4 Inbound, outbound, and in-plant milk runs

Figure 11.5 shows an in-plant milk run plan in a French truck factory.

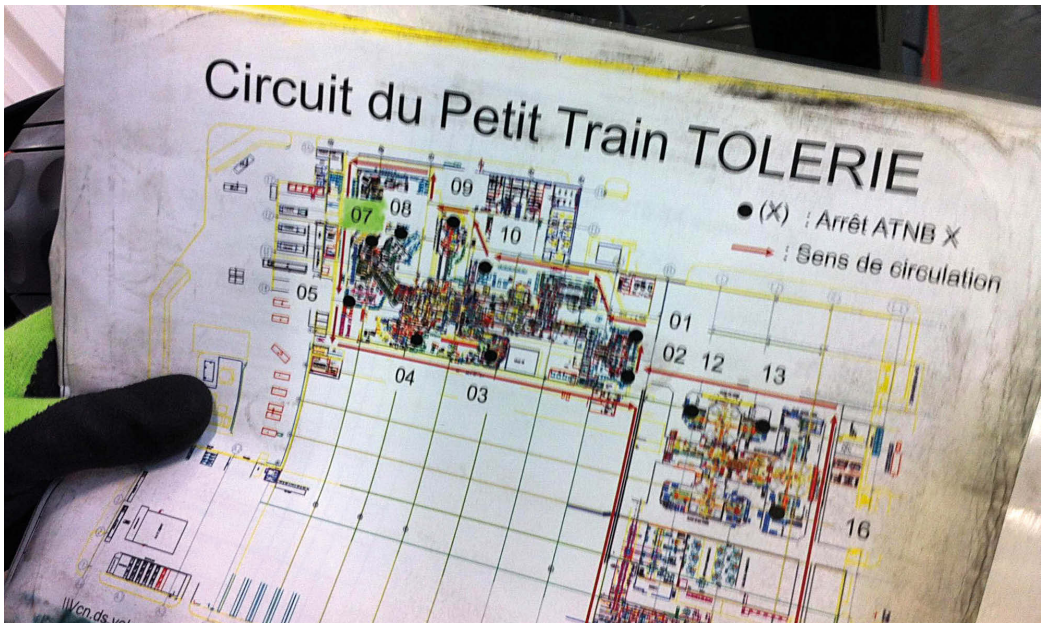


Figure 11.5 The milk run route in a truck factory in France

11.1.5. Using forklifts for transportation

Forklifts moving pallets are a common sight in many plants, and probably the most common method of transportation, both where it is appropriate and where other methods would work better. To use forklifts, you need to answer the following questions:

1. Who orders the forklifts to take action?
2. How is a specific forklift assigned to each task?

Once the decision to move materials has been made, a forklift is assigned to it by a method that is similar to the way taxis were dispatched before Uber disrupted the market. The person requesting a transport calls a central dispatcher who communicates by radio with the fleet of forklifts and sends the next available one. Unless money is no object, however, consumers only use taxis occasionally, to go to the airport, return home from a party, or rush to a hospital, but not for everyday commuting between two locations. In most of the world's cities, they use buses and trains that run at regular intervals along fixed routes, picking up and dropping off passengers at scheduled stops. It is an order of magnitude cheaper, and the duration of the trip is predictable.

The first issue with using a taxi-dispatching model on the shop floor is the time it takes for a forklift to be available. If a taxi company in a city has 100 cars each occupied 90% of the time, then they will almost never be all occupied simultaneously, because the probability of this ever happening is:

$$\text{Probability (all taxis occupied)} = (90\%)100 = 26 \text{ ppm} = 0.0026\%$$

But, if a shop floor has 10 forklifts occupied 90% of the time, then:

$$\text{Probability (all forklifts occupied)} = (90\%)10 = 35\%$$

This means that 1/3 of the time, no forklift is available and transportation must wait. With this level of activity, if we want to have one forklift available 99% of the time, then we need to have 13 forklifts, each one busy 70% of the time. Given the cost of forklifts and the need for qualified drivers, holding an average of four forklifts idle at any time is an unattractive proposition. However, anything less will cause transportation operations to be queued and break the flow of materials.

The second issue is that the forklift and pallet model is a poor fit to the repeatable flow of items between the same locations. Pushcarts or trains of tow-carts traveling on fixed routes at fixed intervals, like city buses and subways, picking up and dropping off quantities of materials that are usually less than full pallet loads provide a cheaper, more reliable, and more predictable service.

This is the in-plant version of the "milk run" concept, and is illustrated in the example in Figure 11.6. The U-shaped cells shown comprise the area of responsibility of one supervisor and have their input and output buffers facing a transportation aisle. The milk run operator from the Materials department deposits full bins and picks up empties on the racks lining the edges of the manufacturing islands. These may be flow racks feeding the cells directly, or the parts may be kitted for assembly by a water spider from the Production department. The milk run operator in this example goes through the figure 8 pattern shown once every 30 minutes.

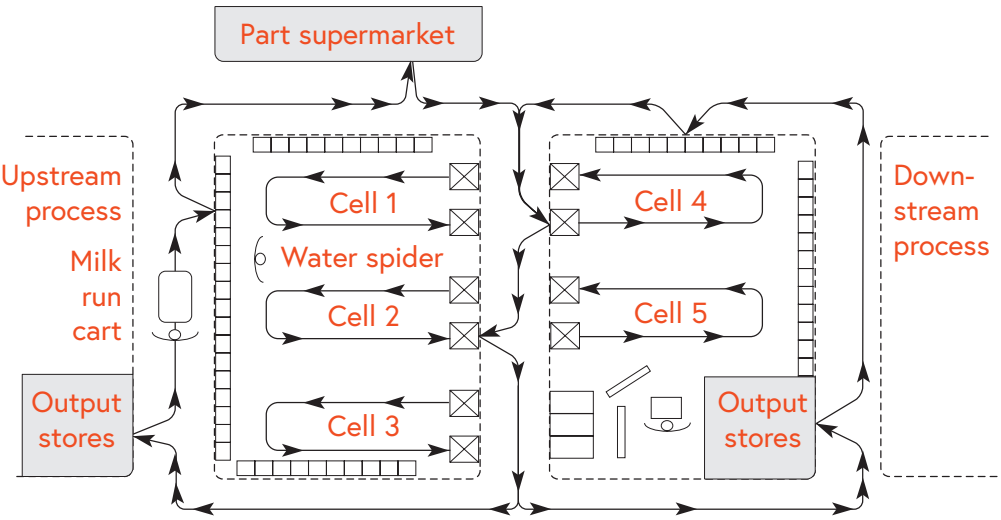


Figure 11.6 In-plant milk run

Compared with supplier milk runs, in-plant milk runs have the following characteristics:

- Their pitch is in tens of minutes rather than hours.
- The pick up and delivery quantities range from single pieces to bins, as opposed to pallets.
- They are managed within a single department of a single company.
- Milk runs in the controlled environment of the shop floor are not affected by weather or traffic conditions, and their schedules are correspondingly more reliable.

There are many possible variations of in-plant milk runs. Figure 11.7 shows a design for low-volume modular assembly of electronic systems. The modules are printed circuit boards assembled and encased in cells. Final assembly consists of inserting the modules into a chassis and closing the chassis with a front panel on one side and a battery compartment on the other. After final assembly, the units undergo

extensive testing, after which many of the units need troubleshooting and repair. While the engineers and managers recognize that "troubleshooting and repair" is not a normal part of a manufacturing process, they are not in a position to get rid of it now, and, by the time they may be able to eliminate this need, the product will have been replaced by a new one, with a process that still needs debugging.

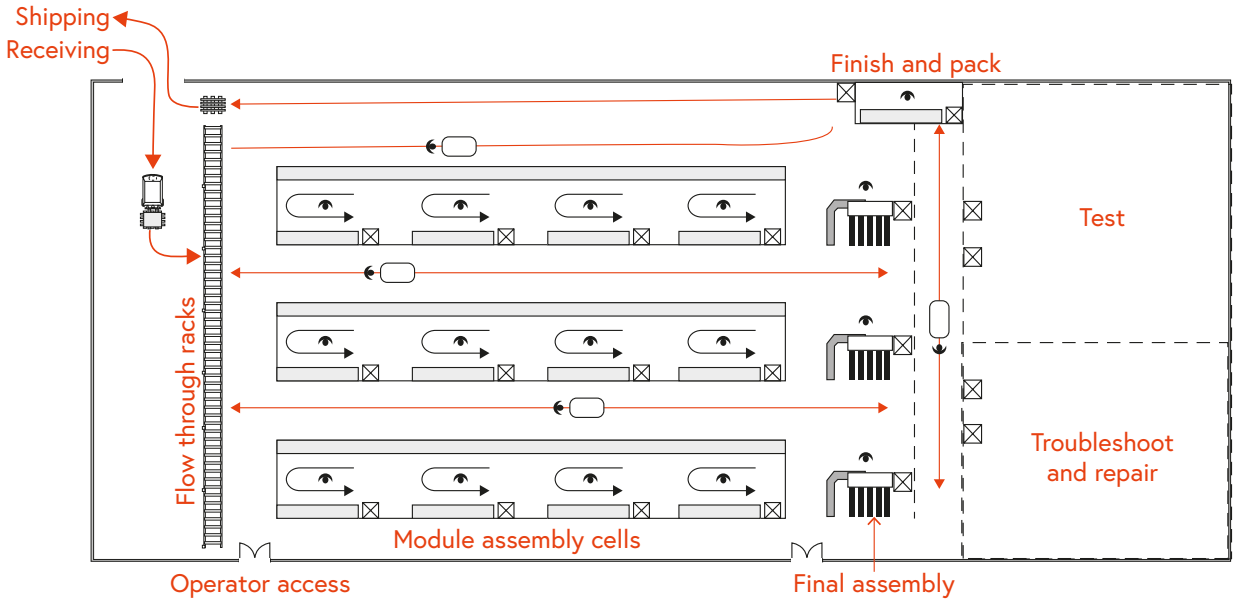


Figure 11.7 Variation in modular electronics assembly

Forklifts deliver parts to a supermarket behind the lines. The need for milk run operators varies with the production volume, and the idea is for one operator to take care of multiple shuttles. Also note the strict separation between areas accessible to forklifts and to people. One forklift delivers parts in weekly quantities to a flow-through rack and takes finished units to Shipping on the opposite side of the rack, milk run operators load parts onto pushcarts that they shuttle up and down the aisles between rows of assembly cells.

Overall, milk run operators move the following:

- Parts from the rack to the module cells
- Finished modules to final assembly
- Assembled units to testing
- Defective units to troubleshoot and repair
- Repaired units to testing
- Good tested units to finish and pack
- Packed units to the output buffer

Box 11.1 **Space attacks**

This box is not about a science fiction movie but about something that happens every day in factories. Space that is unoccupied tend to become occupied for "storage," often of unidentified objects. One way to try to hinder this behavior is to physically bar the space to be protected, as in the attempt shown in Figure 11.8.



Figure 11.8 Barred off shelf spaces to avoid space attacks

However, tape and warning signs are not fool proof. In one project on the shop floor of a household product plant, the team cleared an area destined to hold an assembly cell and marked its boundary with colored tape. Within minutes, a forklift came by and dropped a pallet of unrelated materials inside the reserved boundary. The team stopped the driver and made him take the parts elsewhere. Over the next few hours multiple forklift drivers attempted to occupy this space about every 20 minutes.

These denial of space attacks abated only after the team raised the issue with the materials manager, but, even then, they did not completely stop. Clearly, the forklift drivers were under orders to deliver materials to the floor that had not been requested and for which no storage space was prepared, and the materials management organization had the authority to issue such orders. This is the essence of a "push" system.

11.2. **Water spiders and supermarkets**

A water spider is a person whose main job is to make sure that materials are available in a practical presentation where and when they are needed for production. Hence, a water spider is a job position rather than a method. It serves assemblers on a production assembly line, who are usually organized in teams from 4 to 10 members. The water spider is a fully cross-trained member of an assembly team and can relieve any other members for short periods of time like bathroom breaks. A water spider is not just a "floater" whose only role is to fill in for others. The water spider's special duty is to prepare parts and kits for the other assemblers.

This means unpacking parts if needed and loading them onto such devices as single-piece presentation flow racks or kit pallets. The nickname comes from the way these people dart back and forth between the output buffers of feeder lines, supermarkets, and the input racks of the main line – somewhat similar to how a water spider moves on water (see Box 11.2). The water spiders perform the routine job of picking, unpacking, and kitting parts as needed just before they are used.

Box 11.2 **From "Mizusumashi" to "water spiders" or "water striders"**

If you look up "water spider," you find it is a variety of spiders that hides under the surface of still water and jumps out to capture prey. This is clearly not the right metaphor for the role described here. It's a concept that originated in Japan and is called "mizusumashi" (水すまし), after an insect that dashes back and forth over the water surface. It translates to "whirligig beetle," which doesn't roll off the tongue as well as "water spider." An alternative word used in some lean literature is "water strider," which is an insect that can run on water by using the surface tension and transferring their weight as they walk. Water strider would be a more intuitive translation of the concept, but they are totally different insects than water spiders.

The most common mistake about water spiders is the assumption that this role can be filled by the most junior helper, when in fact it requires the most senior, most versatile operator. While water spiders primarily handle parts, they are not only part of the Production organization, but they are the most expert assemblers. Their relief function obviously makes the skills necessary, but there are other reasons as well. They are more likely to understand the need for and do a good job at preparing parts for assembly than members of the Materials department who may never have put the product together. When kitting parts, in particular, they are less likely to make picking errors.

Figure 11.9 shows an example of a water spider operation. The local stores from which water spiders pick are often called "*supermarkets*." These supermarkets, however, are more like a household refrigerator than a grocery stores. They are small stores of parts located a few tens of feet from the assembly line.

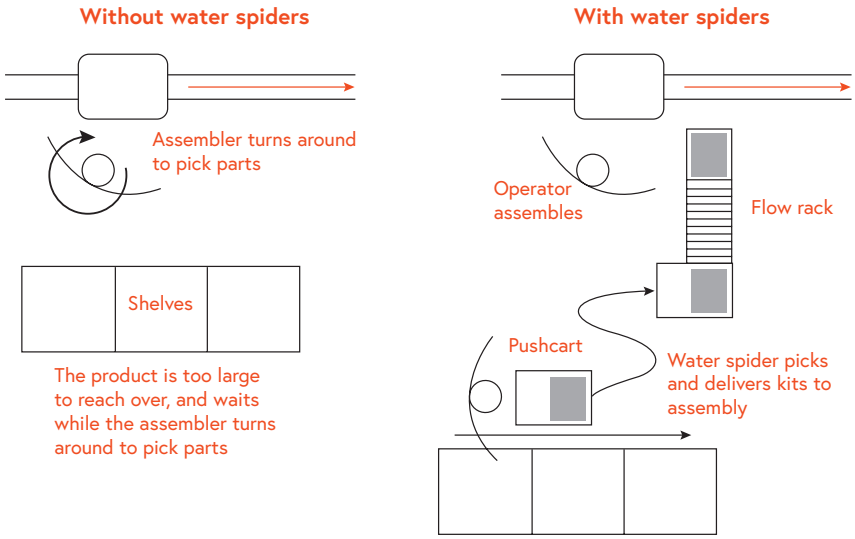


Figure 11.9 Example of water spider operation

According to Taiichi Ohno, the idea of a "supermarket" came to him when a classmate went to the US in 1951–1952 and came back with color photographs of an American supermarket, Piggly-Wiggly. The classmate explained that customers served themselves with what they needed using something looking like a baby stroller and paid at the exit. This was quite different from the Japanese counter system where the customer asked the shop clerk to pick all the things they needed or home delivery sales offering a standard number of household goods. The American system had two major operational benefits:

1. It was much *cheaper* because it *needed fewer clerks* and the *service times were shorter*, and the queues as well. Due to this, they could lower the prices and still be profitable.
2. It *increased sales* as self-service made customers buy things they did not know they needed.

Ohno implemented this idea in the factory in 1952 by having the downstream process go and pick exactly the item needed at the time needed. This is essentially the just-in-time principle.



Figure 11.10 A Piggly-Wiggly supermarket in the 1950s in the USA

Just like in supermarkets, parts arrive into factories in transportation containers that are often not appropriate for use on the lineside. In automotive assembly, for parts that do not arrive in returnable containers, supermarkets are frequently the place where packaging materials are removed and disposed of. In electronics assembly, where the packaging is often much bulkier than the parts, it is sometimes removed at Receiving prior to bringing the parts to supermarkets in returnable bins.

11.2.1. Runners, repeaters, and strangers

In the assembly of custom or highly configurable products, the focus shifts to orders and to the proportion of orders that can be assembled from a subset of all the purchased components. Based on a representative sample of current demand, this analysis sorts the components into "runners" (A-items), "repeaters" (B-items), and "strangers" (C-items) categories, based on objective criteria such as:

- **Runners**: Items without which nothing can be shipped.
- **Strangers**: Items without which most of the orders can be shipped. The threshold can vary.
- **Repeaters**: All other items.

This categorization is then the basis for separate logistics policies. For example, the components could be managed as follows:

- **Runners**: Dedicate easily accessible warehouse locations for each, and load them systematically on milk runs. Consignment and vendor-managed inventory are options for standard items like nuts, bolts, and washers. For product-specific runners, kanbans or the reorder point method may be useful if demand fluctuates.
- **Repeaters**: Dynamically assign storage locations like hotel rooms, and use kanbans to determine when to deliver them through milk runs to supermarkets. Then kit them for delivery to assembly.
- **Strangers**: Don't order from suppliers until you have a customer order for products that use them. Set up a separate storage and retrieval system for these items.

Like P-Q analysis, this categorization and the policies are subject to periodic review, at a frequency that is a function of the company's business environment. This results in promotions and demotions of items between categories, with corresponding changes in logistics. In addition, the analysis is also redone in response to events that are likely to affect the part consumption pattern, such as new product launches or other promotional actions.

Identifying the smallest subset of components with which x% of the orders can be built complete is in principle complicated. The following approximation is, however, tractable:

- Sort and rank the components by decreasing frequency of use. Note that the quantity of interest is not the total volume consumed but the number of orders in which at least one unit is used.
- Count orders that can be built complete with the top n components as follows, working upwards from the least frequently used items:
  - Identify the orders requiring the least frequently used items.
  - Remove these orders from the list.
  - The remaining orders can all be built without these items.

Figure 11.11 shows the resulting curve, which, unlike P-Q charts, is S-shaped. This is understandable, since there is a minimum number of items needed to assemble any order. The figure shows that 80% of the orders could be built complete with the 145 most frequently used components, or 16% of the total of 921 items in the database. To take this analysis further, the same data can identify the 20 parts used in the most orders.

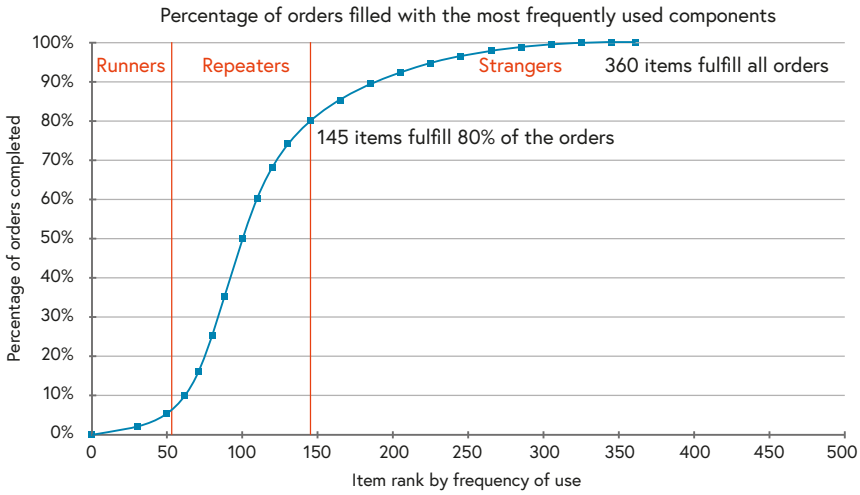


Figure 11.11 Items used in almost every order must be treated differently from those used rarely

How does this numerical analysis translate to regulation mechanisms on the shop floor? The example in Figure 11.12 has a hinged plate requesting replenishment that falls down on the back of the flow rack for the materials handlers to see as they make their rounds. This is best for runners of small items that they routinely carry around on a cart or in a train. If they don't have the item available when they see the plate, then they need to write it down to bring it next time, and a normal kanban would be simpler.

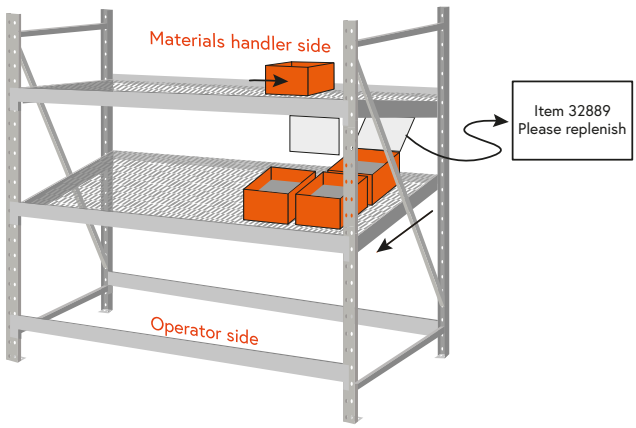


Figure 11.12 Drop-down plate replenishment signal in a simple two-bin system

11.3. **Pull systems**

> **Definition**:
> A *pull system* is one in which the transfer of materials out of one operation is triggered by a signal from the next operation.

A pull is characterized by consumption-based materials replenishment. It is the alternative to push. Pull systems are a counterintuitive way to deliver products faster. Push systems seemingly make products

readily available but they increase WIP, multiply handling steps, and need more handling equipment, ultimately hurting delivery performance more than helping it.

Lean manufacturing is often described as a "pull system," as opposed to the "push system" it replaces. This distinction is only relevant in the logistics domain. It applies to the transportation of parts between plants or between or to lines within a plant. The replenishment logic assumes the immediate future to be like the present, which works for auto parts. Other industries like toys or electronics are subject to explosive growth – during which replenishment can't keep up – followed by plunge in demand – during which it accumulates inventory. In such cases, "dynamic pull" with sophisticated short-term forecasting may be an improvement.

Buying involves decisions, but responding to pull signals doesn't. In the reorder point system, *one single* pull signal is issued for an item when the stock crosses a threshold. This basic system is still used where sufficient. A pull system directs activity without central coordination; this does not work for final assembly and fabrication with complex setups, which require explicit scheduling. Pull signals can be physical tokens or electronic messages. Bins, fixtures, marked spaces, cards, golf balls, and various forms of electronic messages can all be used as pull signals. Before going into the details how they work and where they apply, however, this chapter discusses what they have in common that sets them apart.

11.3.1. How pull systems differ from push systems

The distinction between a pull and a push system is simple (Figure 11.13):

- In a pull system, parts do not move until the destination plant or production line signals that it is ready for them.
- In a push system, parts move as soon as they are ready, regardless of the conditions in the destination plant or line.

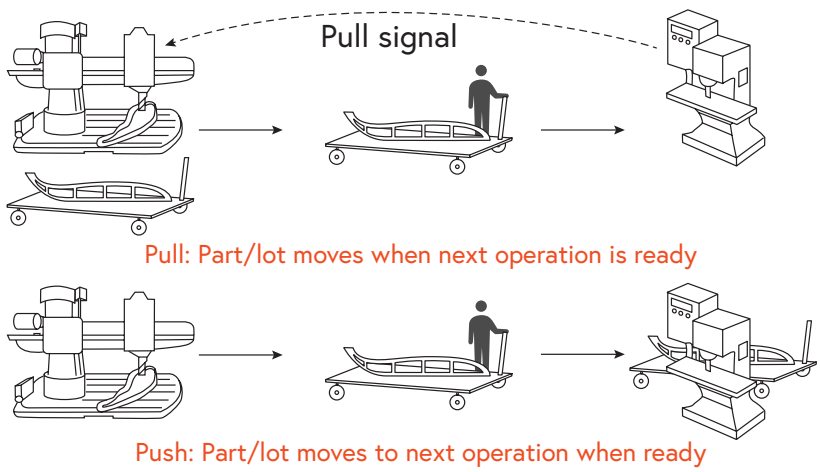


Figure 11.13 Push and pull concepts

In a push system, parts move as soon as *they* are ready; in a pull system, as soon as the *next operation* is ready, as evidenced by a pull signal. Many managers fail to work out the logical consequences of the pull concept. They *want* parts to move on as soon as they are ready, because they assume it will get them

out the door in finished goods faster, but, in a pull system, this is not supposed to happen. Instead, the parts stay in the output buffer of the line until a pull signal arrives, and this output buffer is located right at the end of the line where it is *visible*, not in a separate room. If the line's production runs are for pallets and the pull signals for individual boxes, then boxes are what moves, as shown in Figure 11.14. The output buffer works like a retail store, from which the next line "buys" parts by issuing pull signals, and in which the parts that have not been bought yet remain on the shelves.

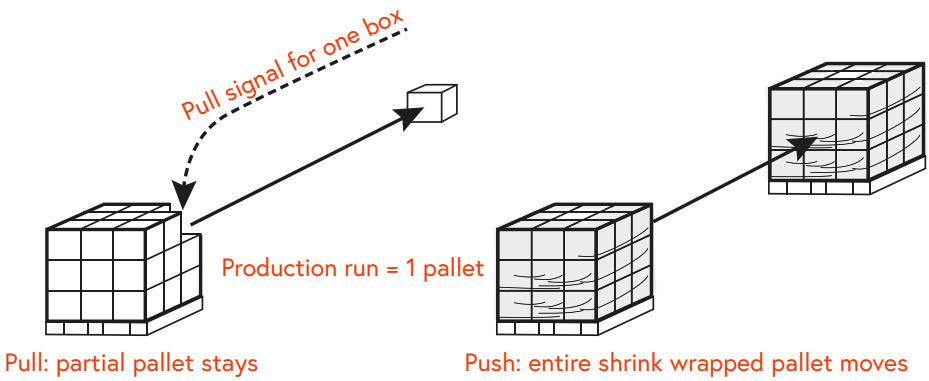


Figure 11.14 Examples of quantities moved in pull versus push system

The move to a pull system on the shop floor is a partial reversal of the trend observed from the 1840s to the 1970s, away from market mechanisms and towards what Alfred Chandler called "the visible hand"[1] of management to allocate resources and work within companies and industries. The attempt to use MRP to plan and schedule the shop floor can be seen as the extreme of the visible hand reaching into operational details and prescribing from a central location which parts should be run and when through each machine.

Pull systems allow local decisions to be made locally, using a logic that makes these decisions naturally consistent with global business needs. The concept is simple, but making it work on the ground with actual products, processes, people, and equipment is not. Pull signals can take on a variety of forms, from empty containers or fixtures to cards called "kanbans" and electronic signals.

In materials handling, *a pull system moves parts between locations when and only when the destination location signals that it is ready for them*. In production, likewise, it feeds parts to the first station of a line when and only when the end of the line signals that it is ready. Sometimes, but not always, the same signal triggers both the production of parts and their movement after completion. In a pull system, parts or lots of parts don't move until the downstream operation is ready to process them.

A complete picture requires an understanding of what "ready" means and the response lead time. The pull signals are not based on anyone's subjective assessment of readiness; instead, they are triggered by part movements or consumption, and are intended to replenish stores at the issuing location up to a specified level. The action taken in response to the signal is not expected to be instantaneous. In lean plants, delivery through milk runs and production with leveled sequencing has predictable lead times, which the pull system relies upon.

1 Chandler Jr., A. D. (1977). *The visible hand: The managerial revolution in American business*. Harvard University Press.

In his book *Toyota Production System*, Taiichi Ohno[2] describes the pull system as being one in which "the down-stream operation comes to take away the materials," as opposed to having the upstream operation bringing them unsolicited. This conjures up the image of the output buffer of the upstream operation being like a retail store where the downstream operation comes to help itself. While this is a powerful image, it should not be taken to mean that an *operator* from the downstream line leaves it to go fetch parts. Between two lines, the issuance of a pull signal causes a *materials handler* to move the parts in a process that is more akin to mail order than to walking to the nearest convenience store.

Stop and Think 11.2
What do you think are some implementation challenges of pull systems?

11.3.2. Pull systems: A counterintuitive approach

Pulling is simple, but counterintuitive, and many plant managers have a difficult time understanding its implications. Lean manufacturing aims to move product units through the production process as fast as possible, and in particular with the least amount of waiting time between operations.

Holding parts in the output buffer of the operation they just completed hardly seems like the way to get them out the door faster. On the contrary, moving them as soon as they are done to the next operation appears to be the way to expedite their completion. "You can't get anything to move" is the most common complaint of production managers about the Materials organization.

Figure 11.15 shows an example of the push approach in action. As soon as a pallet is full of parts, the upstream operator brings it with a pallet jack to a staging area. A materials handler driving a forklift then transfers the pallets to a pallet rack. The same or another driver later transfers the pallet from the rack to the downstream operation.

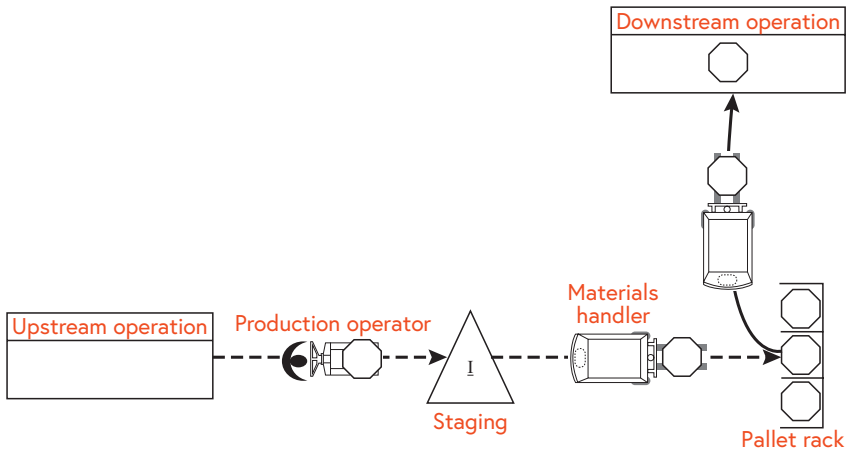


Figure 11.15 A push system

2 Ohno, T. (1988). *Toyota Production System: Beyond large-scale production*. Productivity Press.

In this mode, the parts are transported 3 times and stored twice on their way between the two operations. While this clearly appears wasteful, eliminating it requires asking "why?" several times, as follows:

- Why is there a staging area? Management does not want forklifts in the production area, and therefore production operators need to bring the full pallets to a location where forklifts are allowed. In addition, the output area of the upstream operation is used to stage empty containers and pallets.
- Why do we need forklifts? Pallet jacks cannot store and retrieve pallets in the upper levels of a pallet rack.
- Why do we need a pallet rack? The upstream operation has longer setup times than the downstream operation, and therefore longer production runs. The production run is, say, 6 bins upstream and 1 bin downstream. The pallet rack is needed to hold at least one upstream production run of every item.
- Why stage empty containers near the machine? The materials handler has the time to do it and the space is available.

If the next operation is not ready for the parts, the only effects of moving them are as follows:

1. Shift work-in-process downstream without reducing the production lead-time.
2. Deprive the operators of the upstream operation of information on the downstream operation's ability to absorb their output. The output is out of sight, and they can see neither how much of it there is nor how quickly it is being consumed.
3. Deprive the downstream operators of control over their environment, as materials handlers deliver pallet after pallet of unsolicited materials, overflowing shelves, encroaching on transportation aisles, and reducing visibility by stacking pallets in open areas.

Figure 11.16 shows the same operations connected in a pull system. The space next to the upstream operation is no longer used to store empty containers. In fact, the operation needs at most one empty container in addition to the one currently being filled. The output buffer shown in Figure 11.16 has parallel flow racks, each of which can hold one production run of one item, and the end of the rack is sufficiently low for a pallet to be picked up with a pallet jack. Setup time reduction (SMED) efforts at the upstream operation have made it possible to cut the size of the production runs in half to 3 bins.

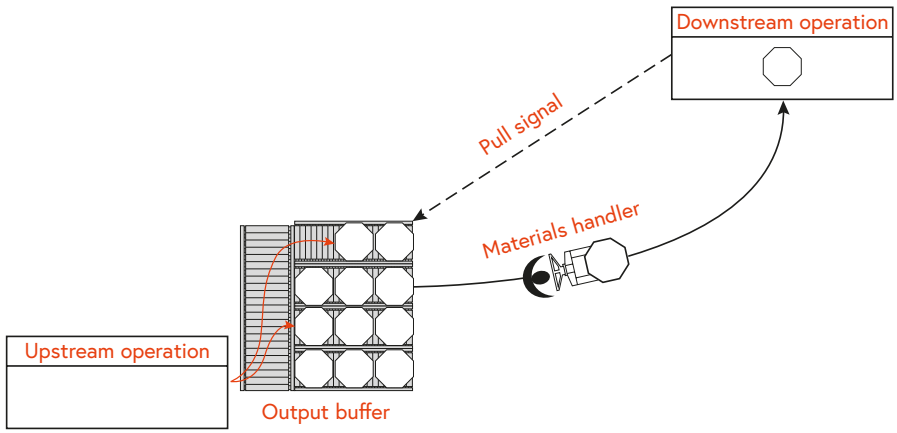


Figure 11.16 Conversion to a pull system

The content of the output buffer is an order of magnitude lower than the content of the staging area and the pallet rack of Figure 11.16, and, as a result, the throughput time of the material is reduced from days to hours. It is possible because the pull signals synchronize production between the upstream and the downstream operation.

We can see why the output of the upstream operation must be held next to it until pulled: It provides visibility for the operator. When the output is held in a separate staging area, as in Figure 11.15, the operator of the upstream operation loses sight of the downstream operation's consumption rate. If, for any reason, the downstream operation falls behind, the upstream operation is likely to overproduce until the staging area and the pallet rack overflow.

Because the amount of WIP between the two operations is limited to the output buffer of the upstream operation, it does not require a multilevel pallet rack, and therefore it needs neither a forklift nor a driver. Note also that, by contrast with Figure 11.15, in Figure 11.16, the pallet jack is pushed by a materials handler instead of the production operator, which allows the production operator to focus on production, as he or she should.

The answers to the "Why?" questions above highlight the implementation challenges. Pulling is faster and cheaper than pushing, but requires more effort in designing the logistics between operations, as well as more discipline in execution.

11.3.3. **Issuing pull signals**

The downstream operation can communicate its needs upstream in many ways. An approach that predates lean manufacturing is the *reorder point system*, in which a replenishment request is issued when the inventory on hand crosses a lower limit. In the reorder point system, the quantity requested is intended to bring the inventory level back up to a maximum quantity. In terms of consumption of the item, the result is shown in Figure 11.17. The cumulative consumption of the item is flat when the line is stopped or making a product that doesn't need it, and rises linearly when it is being consumed per the line's production rate.

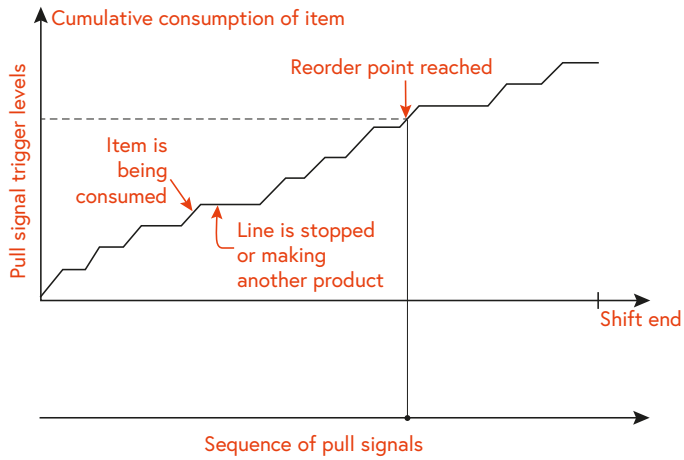


Figure 11.17 Operation of the reorder point system

The reorder point is set at such a level that the remaining inventory is sufficient to cover the demand during the replenishment lead time. Until the reorder point is crossed, no information is communicated back to the upstream operation, and there is at most one signal for the item being acted on at any time. Where these limitations are not a problem, variants of the reorder point system are used in lean manufacturing, but the more common process of issuing pull signals over one shift is as shown in Figure 11.18. Each pull signal is for a fixed quantity, and is issued whenever the cumulative consumption crosses the corresponding levels.

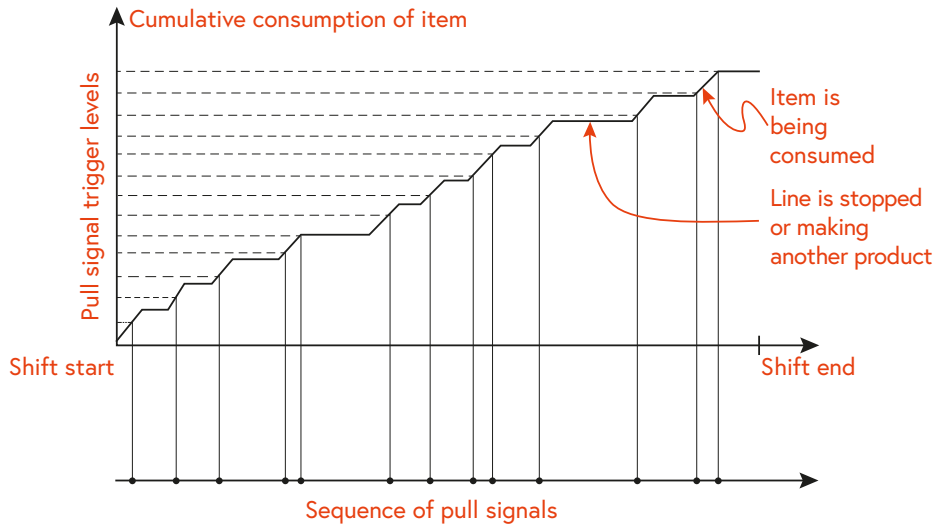


Figure 11.18 Pull signal issuing process

The times between pull signals get shorter or longer in response to variations in the consumption rate. These variations are not an enemy of the pull system; in fact, they are the reason it exists. If there were no variability, and one bin were consumed like clockwork every 60 minutes, then no pull signal would be needed. The kinds of variability the pull system is designed to absorb, however, are as follows:

- Daily fluctuations around a fixed rate, such as are due to trucking times affected by traffic conditions, tool breakage on machines, or operators on short-term sickness leave.
- Small changes in the demand.

There are pull signals for every occasion. In increasing order of complexity, first comes the use of devices that are already on the shop floor and are pressed into service as pull signals through minor modifications. Then comes the use of physical tokens that circulate along the loop, bearing various amounts of data. Finally comes the use of computer, communications and video technology, and cyber-physical systems. Whenever a system is used that depend on human execution, it is essential to make simple protocols (see Box 11.3).

Box 11.3 **Pull signals need simple protocols**

The signals and the protocols for using pull signals must be as simple to execute on the shop floor as possible. Setting them up may require hard thinking on the part of managers and engineers in production, production control and materials management/logistics, but executing them shouldn't. Protocols executed by people are constrained to be simple, as can be seen in the example of a crossroad with four-way stops on US roads, shown in Figure 11.19.

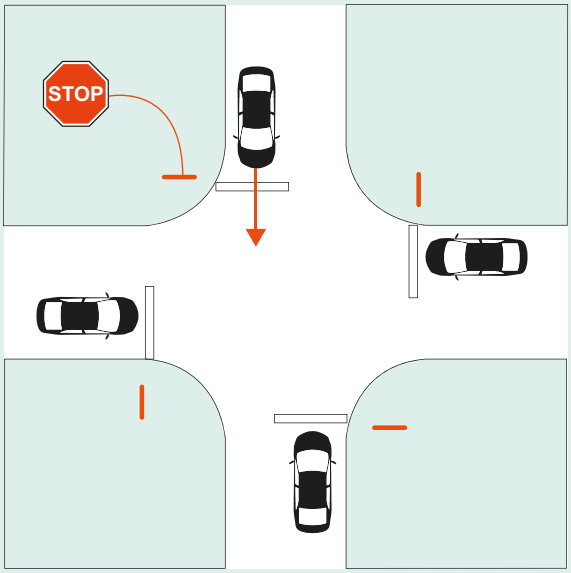


Figure 11.19 The four-way stop protocol

The rule is that the car that has been waiting the longest from any direction gets to cross next. This is as complex a protocol as human beings in large numbers can use, and the only reason it works is that the stakes are high, since violating it could result in fender benders or injuries. Even as simple as it is, there are many countries that find it too dangerous to use.

Imagine what would happen if it were made more complex, for example as follows:

1. A Mercedes has first priority.
2. If there is more than one Mercedes waiting, then the one that has been waiting the longest goes first.
3. If there are no Mercedes, then the first car in the direction with most cars waiting goes first.

It is not difficult to imagine how any attempt to use this protocol would fail, simply because the people involved would get confused or balk at its unfairness. Another sure way to generate chaos would be to have different rules apply at different crossroads.

Computer operating systems and communication networks use much more complicated protocols, but these protocols do not require human intervention, and this is why they can work. Manufacturing logistics is complex, but it is the job of the professionals in the field to reduce it to rules that are easy to follow for operators, material handlers, or truck drivers, and applied consistently. The signals must be easy to see and understand, and they must also have the same meaning wherever used inside the plant or with suppliers. Consistency in the signals from customers is also desirable, but a supplier is rarely in a position to demand it.

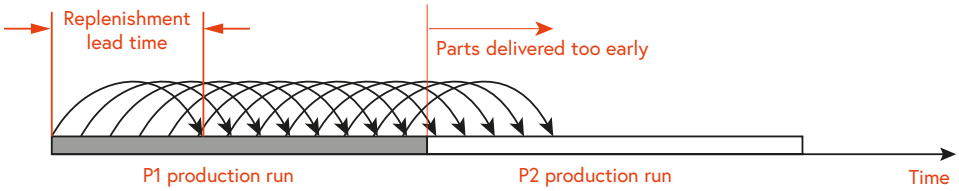
11.3.4. Pull signals in dedicated versus mixed-flow lines

Pull signals are usually not used alone but in conjunction with other systems that forecast activity and generate explicit sequences and schedules for pacemaker lines. The pacemaker is most commonly final assembly, from which pull signals are issued that drive the rest of the plant. Explicit sequencing and scheduling is required in the following circumstances:

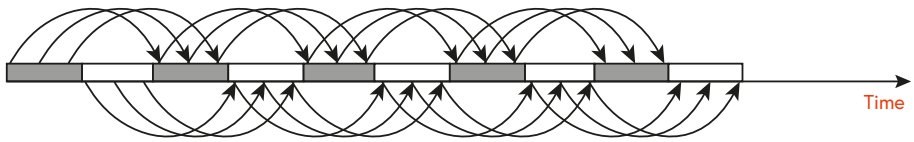
1. The line stops for setups and following the raw sequence of pull signals would require so many setups that the line would not have the capacity to keep up.
2. The raw sequence of pull signals contains random variations in product mix that are better smoothed out before passing the flow of signals onto upstream feeder lines and suppliers.

In a *dedicated* line – a line that always produce one specific product – the applicability of the pull system depends on the stability of the demand. The demand for automobile parts to go into new cars is stable for months at a time, and subject to local, short-term fluctuations due to supply chain dynamics. In this case, we can expect a naïve forecast looking ahead a few hours to be accurate enough for use. On the other hand, the demand for a fashion item sold to teenagers is too erratic, and so is the demand for rarely needed spare parts. The pull system is better suited for automobile production than fashion items.

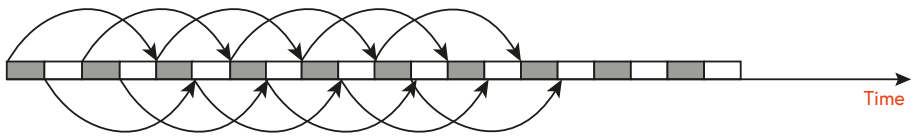
In a *mixed-flow line*, it also depends on the relative length of production runs and the replenishment lead time. For mixed-flow lines, Figure 11.20 shows the simple example of two products, with pull signals that are issued continuously during each run.



a. Production runs longer than the replenishment lead time



b. Production runs at half the replenishment lead time



c. Production runs at 1/4 the replenishment lead time

Figure 11.20 Production runs and replenishment lead times in a mixed-flow line

If the production runs are longer than the replenishment lead times (see Figure 11.20a), then the naive forecast at the end of the run for product P1 is wrong, and parts for P1 are delivered during the

beginning of the run for product P2. This effect is minimal when the replenishment lead time is very short compared to the production run, and worst when they are equal. On the other hand, the shorter production runs are compared to the delivery lead times, the better the forecast and the smoother the delivery pattern (see Figure 11.20b and c). This can of course be accomplished by reducing and level-sequencing production runs, but it can also be done by postponing deliveries. Pull signals can carry parameters indicating that they are for delivery not on the next milk run but on the 2nd, 3rd, or the nth milk run afterwards.

The pull system determines how pull signals are issued and their fulfillment lead time. There are formulas for calculating the number of signals to put into circulation (see, for example, Box 11.4). The idea behind these formulas is to calculate the number of signals that would be needed if production and logistics worked perfectly, and apply a safety factor to compensate for their failure to do so. The major weakness of these formulas is their dependence on an arbitrary safety factor. Its intent is to mitigate the risk of shortage without driving up the amount of downstream inventory any more than necessary. But this risk and the safety requirements can only be assessed based on the consumption pattern, which is not considered in the formulas.

Box 11.4 **Calculating the number of kanbans in a loop**

The formulas in the literature are all variations on Little's law, which says that, in a steady-state, on the average, any input-process-output black box satisfies a simple formula:

$$\text{Inventory} = \text{Throughput} \times \text{Lead time}$$

When pulled from a part or a bin, a kanban enters a process that takes it through a loop. It moves to where it is attached to another part or bin, returns attached, and is pulled again. Here, the parameters of Little's law take on the following meaning:

- The *Inventory* is the number of kanbans in the loop.
- The *Throughput = Demand/Bin quantity* is the rate at which kanbans are pulled.
- The *Lead Time* is the time between two consecutive pulls on the same kanban.

The *Replenishment Lead Time* is how long it takes for a pulled kanban to be delivered attached to a part or bin. Unless *Lead Time* ≥ *Replenishment Lead Time*, the number of kanbans cannot sustain the *Throughput*. How much larger it needs be is a question that Little's law does not answer because it is about *means* and not *fluctuations* around means.

The published formulas pad the *Replenishment Lead Time* with an arbitrary safety factor, say 20%. Milk runs, among other advantages, make replenishment lead times more predictable, mitigating the need for padding. Inside the plant, a small train coming around every 30 minutes can deliver an item every 30, 60, or 90 minutes with minimal fluctuations, taking an unknown out of the calculations.

In general, discrete-event simulations of kanban loops are a more effective way to find an appropriate safety factor, followed by monitoring of each loop's performance.

Box 11.5 **Calculation of pull signals in a pitch system**

Some manufacturers go as far as breaking down the work day into "pitches" corresponding to the minimum time between consecutive milk runs, as shown in Figure 11.21. Clocks on the shop floor display the current pitch number instead of the regular time, replenishment lead times are expressed in pitches, milk run periods are 1 pitch, 2 pitches, 3 pitches, etc., with corresponding delivery quantities.

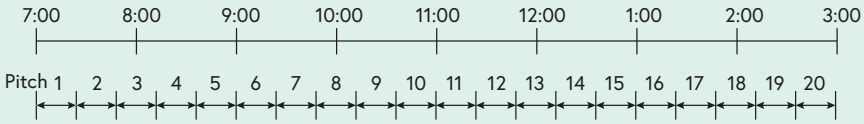


Figure 11.21 Clock time and pitches

While this approach further simplifies calculations, it is not without drawbacks, as it requires training the workforce to reckon time in pitches instead of the familiar hours and minutes, and the organization of breaks is also a challenge. For example, if you want pitches to represent only worked time, then you could stop the pitch clock during breaks, but then it requires the entire plant to take breaks at the same time or else different lines could be at different pitch numbers at the same time.

Using pitches, Little's law, and a safety factor, we can calculate the number of pull signals as follows:

$$\textit{Demand in pull signals per pitch} = \frac{\textit{Demand in bins per pitch}}{\textit{Number of bins per pull signal}}$$

As indicated earlier, the use of a safety factor is a severe limitation of the formula. In fact, its main use is to provide a lower bound to the number of signals in circulation. The most successful implementation strategy is to start out with enough signals to support the usually excessive quantity of WIP initially present in the loop, and then to whittle it down by removing signals one by one until reaching a limit below which shortages begin to occur.

11.4. **Part presentation**

The way parts are presented to operators impacts both productivity and quality. Assembler time is expensive because operators work in series on the line, but materials handlers, on the other hand, work in parallel. The Materials department is the pit crew to the assembler's race car driver, and should focus first on providing effective support to assembly and second on doing it efficiently. Regardless of what managers may believe about "non-value added" activities, the number of people handling and preparing parts is often increased in order to reduce the total amount of labor spent on assembly.

11.4.1. The importance of part presentation

Parts should be presented to assemblers unpacked, within arm's reach, with their smallest dimensions facing out, and oriented for easy installation. Deliveries to the line should be at fixed intervals, in quantities matching the consumption rate, and in returnable containers with dividers or part-specific dunnage, to facilitate counting and prevent errors. The picking strategies of kitting and lineside supply should not be viewed as mutually exclusive but used in combination. When used, kitting should be for single product units. It should be done just before assembly, and into trays with kit-specific dunnage.

Single-piece presentation and water spiders are further techniques used in particular to allow assemblers of large products to pick parts without turning around in "washing machine" movements. It is the last time each part is handled before being mounted on the product, and management often underestimates its impact on productivity and quality. Present the parts too far from the assembly station and in the wrong orientation, and you add seconds of walking and handling time to the assembler's job that constrict the capacity of the entire assembly line and needlessly increase fatigue at the end of the shift. Make the parts easy to confuse and operators will confuse them, resulting in defective products.

In a balanced production line, the assemblers' time is extraordinarily valuable, and therefore we want to relieve them of any work that is not directly assembly. Such work can done by someone who works in parallel with the assemblers. This is why we must be adamant about supplying parts to assembly unpacked, within arm's reach and properly oriented. Conversely, for transporters and materials handlers, the quality of service they provide to assembly is much more important than the efficiency with which they provide it.

Most manufacturing managers are unreceptive to this message. They perceive all materials handling work to be "non-value added," and reinforcing that function is not their objective. Operators pushing carts of parts or transferring parts between containers are typically viewed as non-value added work, because they do not transform the parts in any way that would contribute to what customers want.

In daily life, however, we do not hesitate to pay higher prices for retail than for wholesale, and, by doing so, we recognize the retail markup as the value added to products by moving them and transferring them from pallets and crates to display cases and shelves. It is inconsistent to recognize that this work adds value in the grocery store while denying it on the shop floor. In most plants, the materials manager is measured more on the efficiency of the department as a standalone unit than on its effectiveness in support of production. Changing this around is one of the key motivations for product alignment in the organization structure, where a product line manager has authority over not only the production line itself but its support structure as well.

11.4.2. Key principles of part presentation

The following methods and ideas are applicable in a variety of circumstances:

- Single-piece presentation
- Removal of packaging materials before delivery
- Location within arm's reach of the assembler
- Orientation
- Adjustments to specific part characteristics
- Matching quantities

- Containers with dunnage for counting
- Kitting versus lineside supply.

Single-piece presentation

Single-piece presentation is based on the idea that shelf space facing the assembly line is "oceanfront property" and that everything behind it is less valuable. To put this precious real estate to good use, the single-piece presentation strategy entails the following:

- One and only one unit of each item is presented, with its smallest dimensions facing out.
- Once it is picked, the next unit falls into place automatically.
- Replenishment takes place from behind, without disrupting production.

Single-piece presentation has many advantages:

- It maximizes the number of different items presented at each station.
- It eliminates variability in picking times.
- "Blind picks" are possible. Since an item is always available in the exact same place in the same orientation, where parts are presented behind the assembler, single-piece presentation can eliminate the "washing machine" movements. Once the assembler's hand remembers where the part is, he or she can pick it without turning around.
- Automation is simplified. Since a part is always presented in the same place and in the same orientation, a simple robot can be programmed to pick it.

Once explained, the advantages of single-piece presentation are obvious. It is, however, not a panacea. There are parts, such as wire harnesses, that do not readily lend themselves to this approach.

Removal of packaging materials before delivery

Many parts arrive at the assembly plant protected by packaging that needs to be removed prior to installing the part in a product. Somewhere between the receiving dock and the assembly line, this work must be done. There are many possible ways to organize it. If it is done at Receiving, then disposal of packaging materials is the easiest, but containers have to be provided to hold the parts from there to the line side. On the other hand, if it is done just behind the line, then the collection of used packaging materials needs to be organized plant-wide. The one person who should not be unpacking, however, is the one who is in most assembly plants; the assembler.

While the assembler fusses with packaging materials, the product waits, and it shouldn't. Asking the assembler to do this is like asking a race car driver to change tires or pump gas. Regardless of where and how it is accomplished, when parts are presented to assemblers on the line, they should be in such a condition that they can be picked up and installed in the product directly. Finding the way to do this is sometimes challenging with parts that are easily damaged, such as windshields in cars or electronic components.

Brittle glass parts or scratch-prone painted parts need foam-padded racks and separators. Printed circuit boards and integrated circuits need special dunnage for protection against electrostatic discharge (ESD), or need to be kept in a Faraday cage. These frequently cited reasons for keeping parts packaged until used can be overcome in two ways:

1 By using special containers or bins designed specifically for this purpose. These are known problems, to which many suppliers offer solutions. Figure 11.22 shows the example of a printed circuit board.
2 By limiting the exposure time. Parts need to be out of their protective packages when assembled anyway, and our only constraint is that the assembler not be burdened with unpacking. If they are unpacked just behind the line a few minutes before they are used, the opportunity for damage is limited in space and time and is essentially the same as if the assembler did the unpacking. Figure 11.23 shows an example of how this might be organized in some cases.

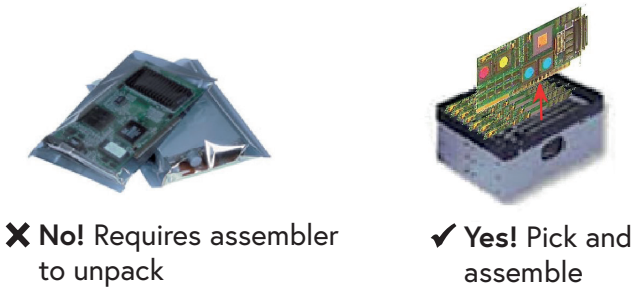


Figure 11.22 Printed circuit boards with ESD protection

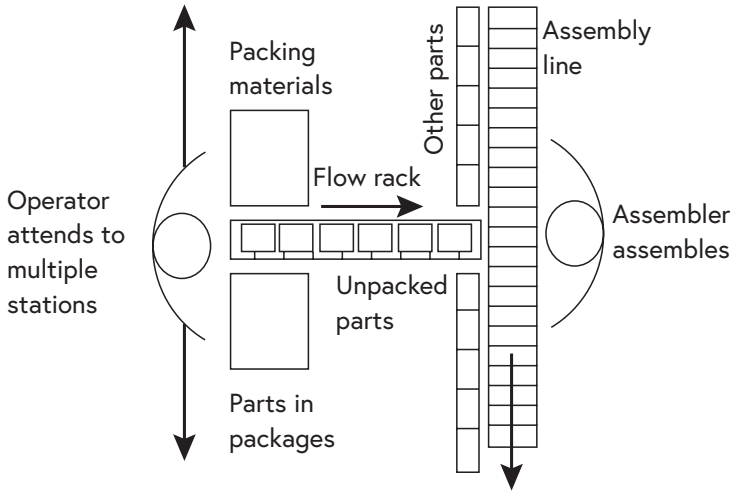


Figure 11.23 Separation of unpacking from assembly

Location within arm's reach of the assembler

What many companies call "point of use" storage is in fact tens of feet away from assembly stations and requires operators to walk back and forth. In others, parts are just far enough to require the operator to turn around for picking in the "washing machine" movement. Assemblers then naturally try to make the most of each turn by picking several items each time. They then do one of the following:

- Keep them in one hand while assembling with the other. This slows them down, as they should be able to assemble with both hands.
- Place them in their pockets or on the assembly station, which is multiple handling.

To avoid this, we must find a way to present parts within arm's reach of the assembler. The following two methods can be used for this purpose: The most obvious way is to have them move down a flow rack to the station, as shown in Figure 11.24. An alternative is to make assembly fixtures double as kit pallets and have some parts travel along with the product units down the assembly line.



Figure 11.24 Gravity flow racks

Orientation

The attention warranted by the orientation of a part depends on its size, shape, weight, and frequency of use. Following are a few extreme examples of parts with different orientation requirements:

- The balls that went into computer mice or bearings are orientation-free. More generally, the more symmetries a part has, the less of an issue orientation is.
- For screws and other fasteners, volume is the key factor. They are most commonly provided randomly oriented in bins. In high-volume operations, automatic feeders are used to deliver them in a controlled orientation.
- Some airliner parts are tens of feet long but less than 1 foot thick and weigh thousands of pounds. Even at takt times measured in days, you want to present them to assembly in such a way that you just have to slide them in, as opposed to lifting and rotating them.

Adjustments to specific part characteristics

Presentation methods must always be adapted to the physical characteristics of the parts. In the orientation section, we have already discussed the case of long and narrow parts. Other categories needing special treatment include flexible parts like wire harnesses and tangle-prone parts like springs.

Pallets are in general inappropriate on the line side, because picking from the near corner on a full pallet does not take the same amount of time as walking to the far corner, squatting, and picking from a near-empty pallet. The problem of picking from pallets can be alleviated by using "pallet pals" (see Figure 11.25); rotating tables mounted on a scissor-lift with springs. Pallet pals allow operators to turn the pallet to the current picking position, and the springs push the pallet pal higher as the load becomes lighter. It is an adequate solution when, for example, one pallet load barely represents 30 minutes of production on items that are easy to carry, as in cookware assembly.



Figure 11.25 A "Pallet pal" adjusts pallet orientation and height

Matching quantities

In many plants, the materials organization delivers items to the line in quantities that occupy the same volume rather than quantities that will support assembly for the same amount of time. Whether it is because Materials uses forklifts and drivers won't waste their time on partial pallets, or because the Automated Storage and Retrieval System (AS/RS) cannot handle partial pallets, the result is that full pallets arrive at the line side, whether one pallet will cover demand for 30 minutes or a week.

If one unit of item X has the volume of 5 units of item Y, then item X is delivered in quantities that are 1/5 of those of item Y, regardless of the numbers consumed. The main consequences are as follows:

1. Without knowing the consumption rates of hundreds of items, a supervisor or an operator cannot see whether, say, a half-empty rack signals an imminent shortage.
2. The different replenishment periods for hundreds or thousands of items combine into a workload for transporters that is characterized by random peaks separated by intervals of idleness, as shown in Figure 11.26.

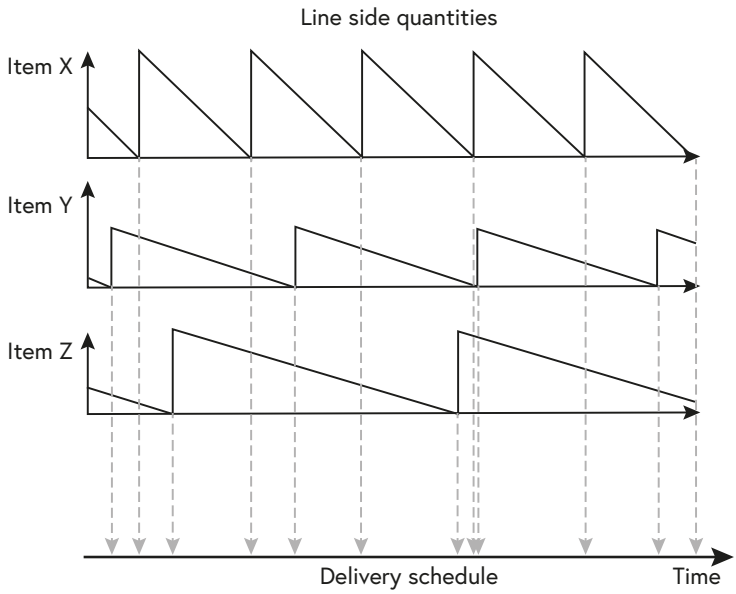


Figure 11.26 Random replenishment schedule

The variability in the aggregate delivery schedule in Figure 11.26 is self-inflicted, because it occurs in spite of constant consumption rates for all items. Its cause is the policy of using forklifts to bring full, single-item pallets to the line. By forcing Materials to cope with artificial peaks and valleys of activity, this policy works *against its own goal* of efficient use of transportation resources.

The quantities delivered to the line should instead be based on consumption rates and calculated to support the line for the same amount of time, or, occasionally, for small multiples of this time. This avoids imbalances in lineside supplies and allows the line to be supported by regular, periodic deliveries of multiple items in matching quantities. This is the "milk run" system described earlier. The delivery vehicles used in implementing that policy tend to be pushcarts or small trains rather than forklifts.

There are items, typically nuts, bolts, washers, or screws, for which one small bin is sufficient to support the line through multiple delivery periods but the delivery period should be set so that all major components are delivered in matching quantities each time.

Containers with dunnage for counting

On many assembly lines, supervisors are first aware of shortages when they actually occur. There are many causes for this, some of which are related to part presentation. Mismatched delivery quantities, as discussed above, are one of them, but lack of part count visibility on the lineside is another. Shortages are intolerable for obvious reasons, but overages are also undesirable because excess parts require attention, are vulnerable to handling damage, and are at risk of obsolescence.

When parts are heaped in bins or in hoppers, it is practically impossible to know how many are there. When spot checking part counts on assembly lines, it is common to find that a box supposed to hold 10 parts actually holds 9 or 12. Part counting on the line side must be supported by structured containers and discipline. Structured containers have slots for individual parts. If an orderly picking discipline is maintained, it is always easy to see how many are there, as shown in Figure 11.27.

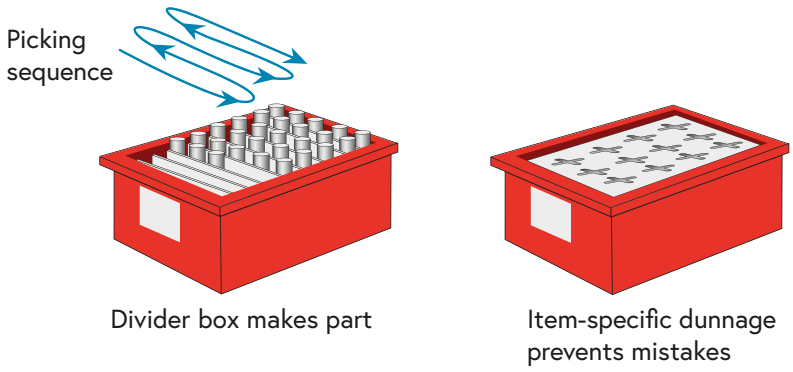


Figure 11.27 Divider boxes and item-specific dunnage

Kitting versus lineside supply

Kitting and lineside supply are two major parts supply strategies that can be used in combination, but usually are not. Some plants are designed around kitting all parts prior to assembly; others, around bringing them individually to lineside stores at assembly stations. In reality, the better strategy is to combine both approaches on the same line.

Kitting is common but most factories' concept of kitting is throwing together all the parts needed to make a *lot* of a product into a tote, in their original packages, sometimes weeks in advance of building the lot. By doing this we commit parts to a particular use before we have to, and we have to "break kits" if, as commonly happens, another use takes priority in the meantime. In short, nothing good and plenty of mishaps can happen to these kits from the time they are picked to the time they are used. The throwing-together-in-a-tote approach also makes the completeness of a kit difficult to assess. Finally, as stated before, unpacking parts is not a job assemblers should do. The proper way to kit is:

- For a single unit of product
- Just before assembly
- With parts unpacked
- Onto a structured pallet with an assigned location for every part

An extreme form of kitting is shown in Figure 11.28. The photo shows a carousel of rings and gaskets ready to be assembled into an automotive part. The kitted parts are to be assembled in the sequence indicated by the red line from upper left to bottom right. This form of kitting is helping the assembler not only pick the right part, but also when to do so. The benefits of this approach has to be weighted against the time investment at the kitting station.

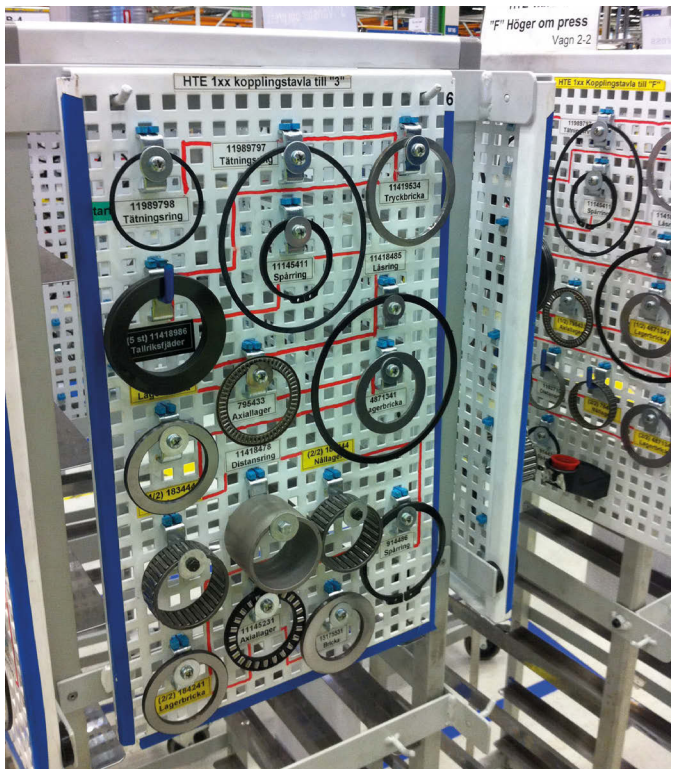


Figure 11.28 In-sequence kitting of rings and gaskets for an assembly operation

The alternative to kitting is to bring all parts to racks on the lineside, for assemblers to pick when they are needed. While it uses more space on the line than kitting, this method has the advantage of requiring less handling and being better able to cope with the presence of defectives. If you pick a defective part on the lineside, you can place it in the reject bin and pick a new one, which you can't do with kits.

A reasonable strategy in a mixed-flow line is lineside supply for all common parts, and for as many product-specific parts as we can fit on the shelves, and kitting for remaining product-specific parts.

> **Stop and Think 11.3**
> How do you organize ingredients and tools when you cook? Do you "kit" everything at the start or search for items as needed? How is that different from assembling a product on the assembly line?

11.5. Logistics technology

Most factories do not take advantage of the full range of logistics technology available to meet their various needs. For example, they may use only forklifts for all tasks (Figure 11.29), which is unnecessarily expensive and gives inadequate service where this device is a poor fit. The following sections review the most common options and their range of applicability.

Inside the plant, structuring the transportation system is the key to many improvements. The principle illustrated in Figure 11.29 may be obvious, but it is routinely violated in factories. Each transportation task should be carried out using methods that are tailored to its load and its frequency. Just as, if you may use the railroad between towns, a car inside the town and a dolly inside the house, you should use appropriate vehicles and an appropriate network of transportation aisles within the plant. The example in Box 11.6 shows what is at stake in this simple idea.

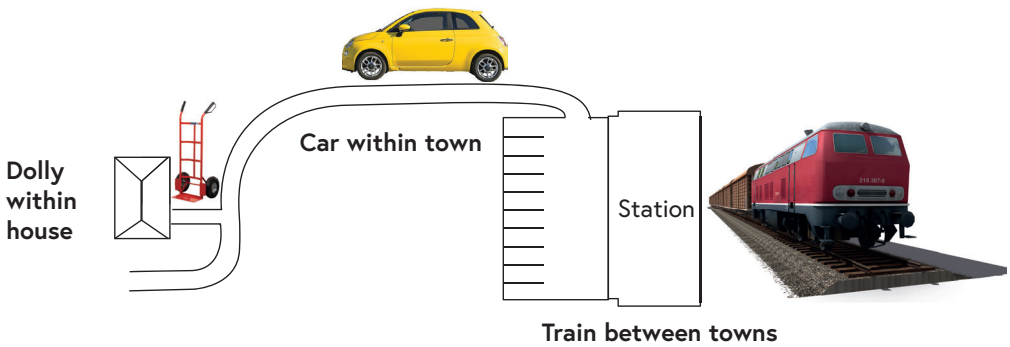


Figure 11.29 A transportation method for each need

Box 11.6 **Improving packaging of dried foods with new transportation system**

This example shows how structuring the transportation system can allow a plant to increase productivity, reduce floor space requirements and improve safety at the same time. Figure 11.30 shows packaging lines for dried foods running at a takt time on the order of 1 sec. The filler fills and seals paper pouches, which the following operations aggregate into larger and larger groups, resulting in a shrink-wrapped pallet ready to load onto a truck. This type of line is fully automatic and the primary job of operators is to keep the machine hoppers full of packaging materials. The machines occasionally jam, and the operators are also expected to clear jams and restart the lines.

In the "before" condition, forklifts brought full pallets of packaging materials right between the lines. In fact, there were no restrictions at all as to where forklifts might go, and no marked transportation aisles. This forced the lines too far apart for one operator to attend to two of them. In addition, the coexistence of people and forklifts in the same environment was uneasy and unsafe, with operators being constantly on the lookout for forklifts.

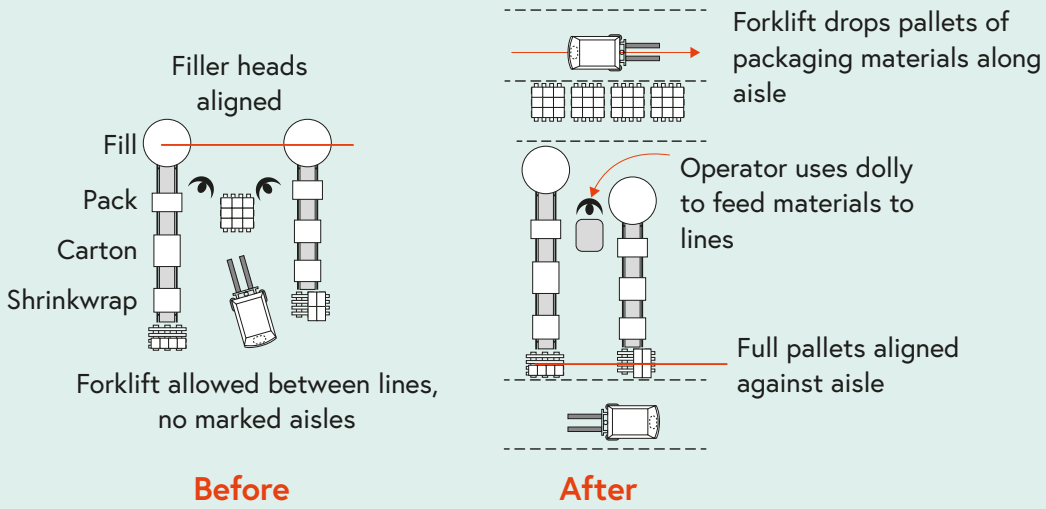


Figure 11.30 Example of packing lines for dried foods

This surfaced when a sister plant closed and this plant was asked to wedge in 5 more packaging lines within the same hall. There was no way density could be increased in the current mode of operations. On the other hand, it could if we did the following: Constrain forklifts to marked transportation aisles.

Assign drop-off and pickup locations along the transportation aisles. One result was aligning the ends of the lines rather than their beginnings. Staggering the fillers actually helped locate more lines in the same floor space, while lining up the full pallets along the transportation aisles.

11.5.1. **Conveyor systems**

Once we have work stations, we need to convey parts from one to the next. There are many options, ranging from manual carrying, via forklifts, to fully automated conveyor systems.

Unpowered conveyance

Examples of unpowered conveyance systems are shown in Figure 11.31. They do not enforce takt time, but neither do powered systems when transfer to the next station is triggered by the operator. When the conveyance system does not enforce takt, then something else has to, and usually it is a production monitoring system.

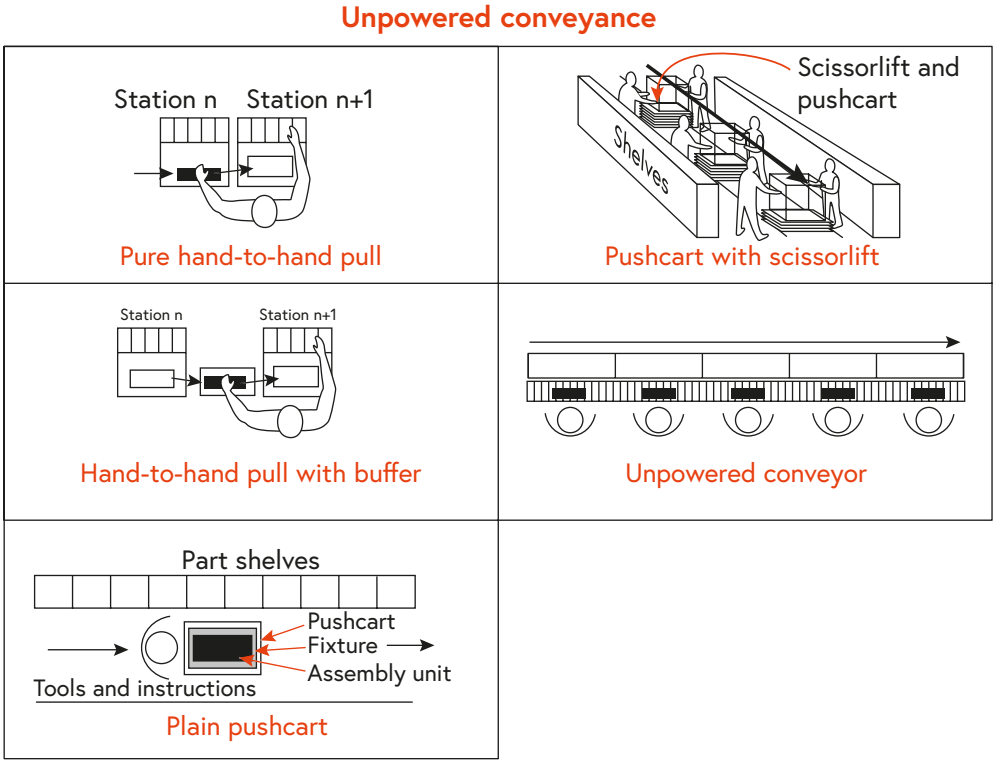


Figure 11.31 Examples of unpowered conveyance systems

The simplest conveyance system is pure hand-to-hand pull, where the downstream operator simply picks up the part the upstream operator just finished. If there is too much interference between the 2 operators, then a 1-piece buffer between the 2 can be set up.

When either the product units or the product unit-fixture combination is too heavy to be passed by hand and volume is low, then a pushcart can act as both a conveyance mechanism and a rolling assembly station. The pushcart with scissor-lift is a variation on this theme, applicable for products that are cubic in shape and filled from the bottom up with components. Such products include large office copiers and semiconductor production equipment. The scissor-lift is at its highest at the first operation and

gradually comes down, allowing assemblers to work at a constant height as the frame fills up. For safety, the scissor-lift is normally skirted.

As volume goes up, so does the value of working on a conveyor. A conveyor does not have to be powered. It may simply have rollers on which assemblers push pallets.

Powered conveyance

Examples of powered conveyances systems are shown in Figure 11.32. Among powered conveyors, we have continuously moving and stop/start types. With continuous motion, the relative position of the work and the assembler changes constantly but work can be done during the entire takt interval. With stop/start conveyors, the work does not move, but some time must be set aside for transfers. The two can coexist in different sections of the same line. Automatic assembly equipment usually requires the stop/start type.

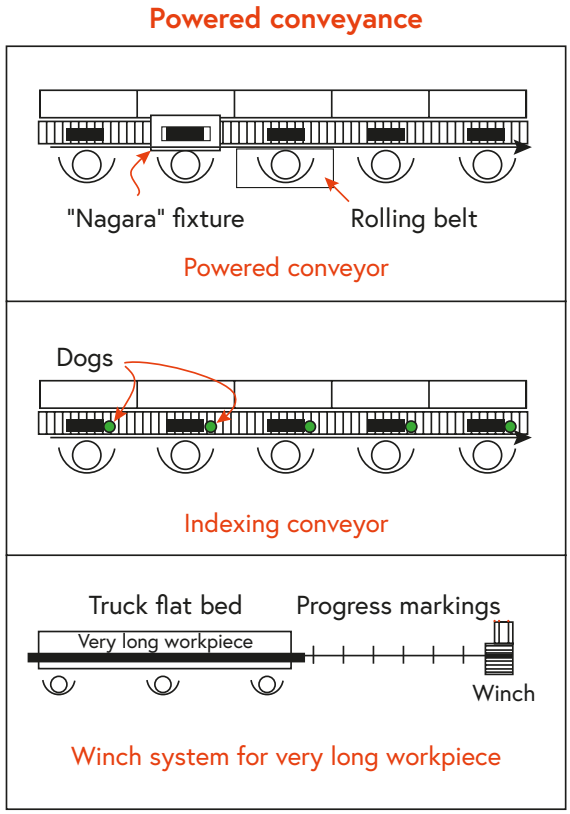


Figure 11.32 Examples of powered conveyance systems

Very long work pieces can be placed on a truck flat bed and pulled forward slowly by a winch. The movement allows operators to work on different sections of the part over time. Markings on the floor track progress. When the part is done, the truck flat bed is hitched to a cab.

Networks of conveyors

Networks of powered conveyors play a central role in many automation projects. They offer a model of continuous flow in small quantities as opposed to the accumulation of discrete loads for transportation by vehicles. Figure 11.33 shows a typical factory conveyor.

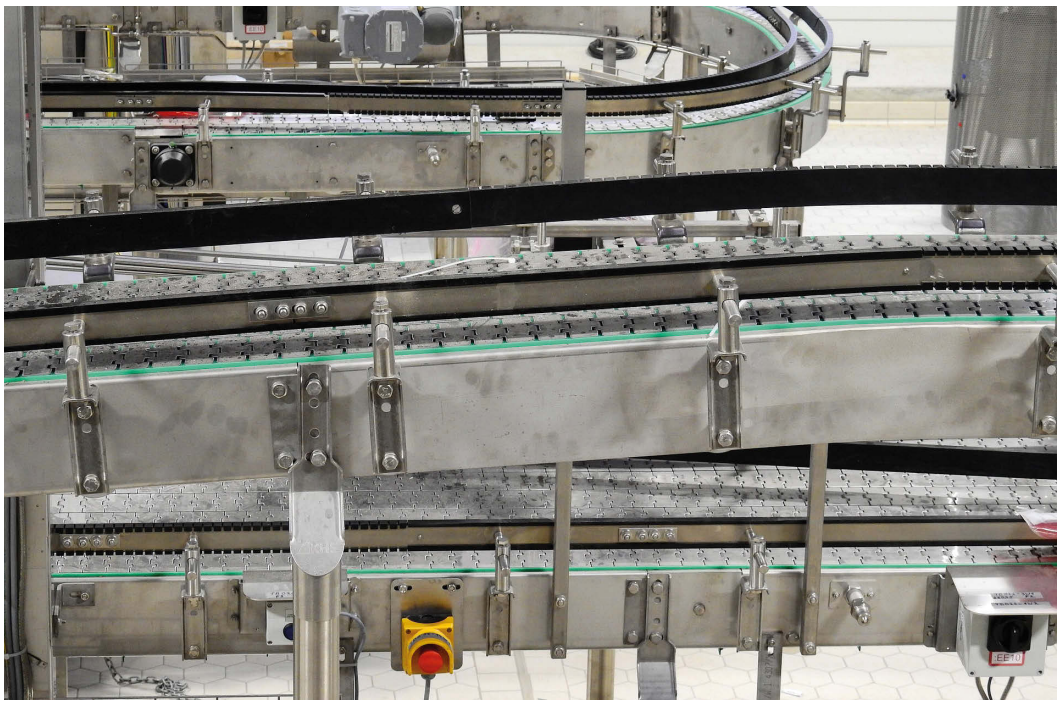


Figure 11.33 Conveyor network

Conveyors are not as prevalent in lean manufacturing plants but also not excluded: They are used more for transferring parts *within* a line than between production lines. If a transportation route between production lines has a sufficient high and stable volume to justify a conveyor between them, they should be collocated with the conveyor correspondingly shortened and made a part of the resulting integrated line. Dell uses conveyors with automatic controls to move kits from picking to assembly and systems to testing and boxing. As these operations are contiguous, it can be argued that this is in fact movement within the production line.

Belt conveyors in various forms and other devices that work like pipelines are commonly used also to move bulk materials, such as foundry sand, pellets of resins in plastic molding or molten aluminum in high-volume die-casting shops. Provided the consumption of the materials is high and stable enough, this method is sound.

Using a network of conveyors to automate palletizing

Many factories have conveyors after final assembly and packaging. Logistics systems for moving standard-size packages are readily available. Figure 11.34 shows two options for palletizing boxes and conveying them to shipping in a high-volume-low-mix plant making products like detergents or toothpaste. Retail units are filled at a rate higher than one per second, and aggregated in multiple stages

into supermarket display cases and cardboard shipping boxes, which are then themselves loaded onto pallets in configurations that vary with the shape of the box.

Figure 11.34a shows this done with no human intervention, using a central automatic palletizer, to which boxes come through a network of conveyors. Figure 11.34b, with one palletizing station for each line and manual loading assisted with devices to lift and rotate pallets and manipulators with suction cups to relieve operators from the weight of boxes.

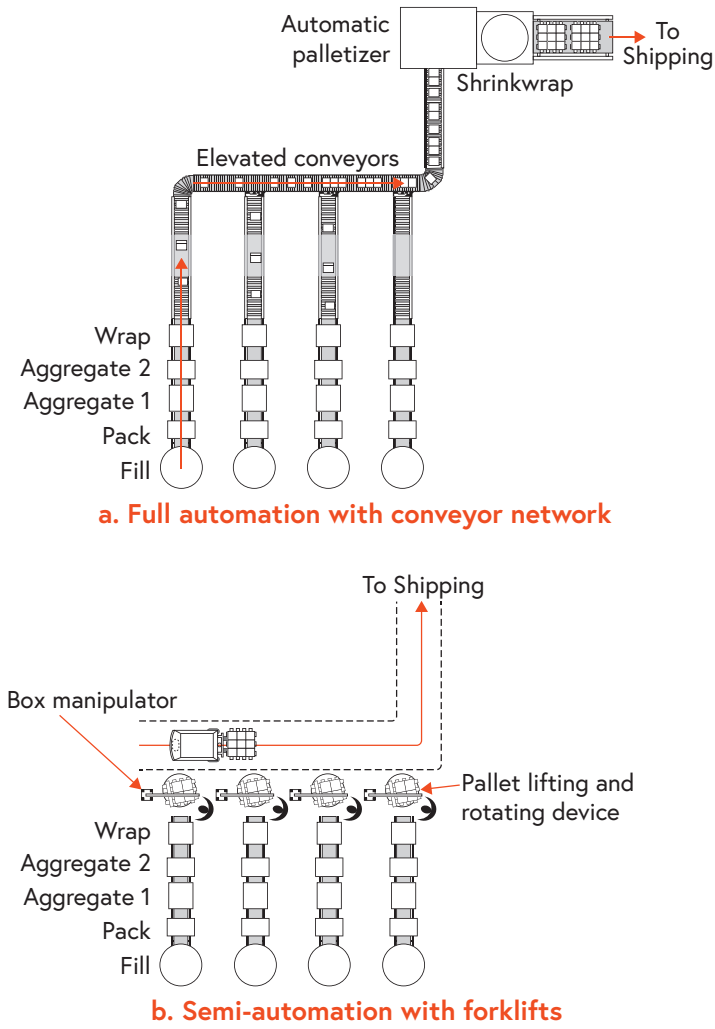


Figure 11.34 Conveyor networks versus semi-automation

Both solutions *a* and *b* are technically feasible and are in use in different plants. Full automation, in principle, should save labor, reduce inventory and eliminate human error. In reality, however, it raises the following concerns:

- It is an order of magnitude more expensive to implement.
- While it saves labor in forklift operations, the labor savings in palletizing may not materialize. Operators are needed to feed the machines along the lines with packaging materials anyway, and the elimination of manual palletizing may not reduce their number.

- Automation adds labor in maintenance and in control system programming and configuration. Once programmed correctly, it should palletize all different types of boxes correctly and reliably. The debugging of this software, however, is usually expensive and drawn out, and so is the process for changing or enhancing it as new needs arise.
- Reliance on one single automatic palletizer and one conveyor network makes the entire plant vulnerable to failure in either. In such plants, the core technology is compounding, mixing, or spray drying, and the automated system diverts engineering attention away from these processes.
- Layout changes are more difficult than with the manual system, as there is more to moving conveyors than to changing aisle marks on the floor.

The approach shown in Figure 11.34a avoids creating such "monuments" by keeping palletizing decentralized, and the use of various devices to make manual palletizing easier, reducing the risks associated with operator fatigue. On the other hand, this approach relies on operators to know the pallet configurations applicable to boxes of varying sizes. This means that they must be trained on it and that human error *will* occur.

It is impossible here to say that one solution is always better than the other. A rational choice can only be made for a plant on a case-by-case basis, considering such issues as labor costs, the skill level of its workforce, and the availability of engineering support services in its neighborhood. Figure 11.34b is consistent with jidoka, which stresses incremental steps focused on making the work easier for people to perform consistently, eventually leading to full automation.

Best practice in conveyance

Following are some of the best practices and corresponding common mistakes in the area of conveyance systems:

- **Conveyors should not be WIP buffers.** The ideal is one-piece flow. Some companies deliberately extend the path between stations by raising the parts and dropping them onto a downward spiral flow rack. The legitimate use for this technology is in coolers or freezers, but, in some plants, they are simply used to accumulate WIP between stations to protect the downstream station against failures in upstream equipment.
- **Conveyors should not separate upstream and downstream stations.** When straight conveyors are used as WIP buffers, they pull upstream and downstream work stations so far apart that operators cannot help or relieve each other.
- **The conveyance system should not lose part orientation.** It is common that a part coming out of a machine in a controlled orientation will be dropped onto a conveyor belt in a random orientation, and that money will be spent at a later stage to recover that orientation.
- **If a conveyor is used, assembly work should be done on it.** Another mistake is to use a conveyor but to have assemblers take parts away from the conveyor to work on them and put them back on afterwards. If you're using a conveyor, then the assembly work should take place on it.
- **Avoid cranes or having assembly units hang on chains.** These devices do not accurately control positioning and orientation. They not only require operators to manually position dangling workpieces but are also unsafe.
- **Avoid shared conveyance systems.** If, in spite of the drawbacks, you use a crane, you don't want to share it among multiple cells, because it makes the feasibility of a move within one cell contingent on what happens in other cells.

- **First-in-first-out (FIFO).** The conveyance system should preserve the process sequence. While FIFO is often perceived as "fair," the real reason to pursue it is quality. Any conveyance system that shuffles the parts reduces process transparency and makes quality problems more difficult to trace. This excludes hoppers, bowl feeders, or even stacks, which are last-in-first-out devices.
- **Takt time enforcement.** In moving assembly lines, takt time is enforced by the movement of the conveyor. There are many types of conveyance systems that do not enforce the pace. This is acceptable, as long as other devices do. In automatic lines with takt times of 5 seconds or less, sensors can be used to draw the control system's attention to takt time violations that are too small for people to detect, yet hurt the performance of the line.
- **Dedication.** The conveyance system should be dedicated to the pair of stations between which it moves parts.
- **Preservation of part orientation.** A conveyance system which keeps the product in a controlled orientation is superior to one that doesn't, and this orientation may not be the same at two subsequent stations. The unit, for example, may need to be rotated 180 degrees, and this is better done while it is moving between stations than after it arrives.
- **Unobtrusiveness.** The conveyance system should be in the background, rendering a valuable service without drawing attention to itself, allowing assemblers to focus on assembly. If all a visitor sees when entering the shop floor is a complex network of conveyors and elevators, the conveyance system fails this particular test. The conveyance system must not interfere with the work and not dwarf the actual assembly stations.
- **Safety.** In an assembly plant, there may be more safety issues with conveyance than with anything else.
- **Low cost of acquisition**, operation, configuration change, and maintenance. Conveyor networks are not only expensive to buy, install, operate, and maintain, but they are also inflexible.

11.5.2. Forklifts and pallet jacks

Forklifts are such common devices that they immediately come to mind in any discussion of in-plant transportation. They are versatile and powerful, but they are not without disadvantages:

1. They cost tens of thousands of dollars.
2. They can only be operated by specially trained drivers.
3. They are a safety hazard and must be constrained to run in designated areas.
4. They are appropriate for moving pallet-sized loads but not smaller quantities.

The common forklift (see Figure 11.35) that travels throughout the plant has relatives like narrow aisle trucks that are specialized for storage and retrieval or picking operations in warehouses, but these are not *transportation* vehicles per se. Forklifts are generally overused in factories.



Figure 11.35 A standard forklift for carrying pallet-sized loads

The manual and powered pallet jacks shown in Figure 11.36 are 1 to 2 orders of magnitude cheaper alternative to forklifts for short, *horizontal* movements and don't require special training. They are not commonly used for long carries, but the only job they absolutely cannot do is access stacks of pallets or the upper shelves of pallet racks or on top of pallet stacks.

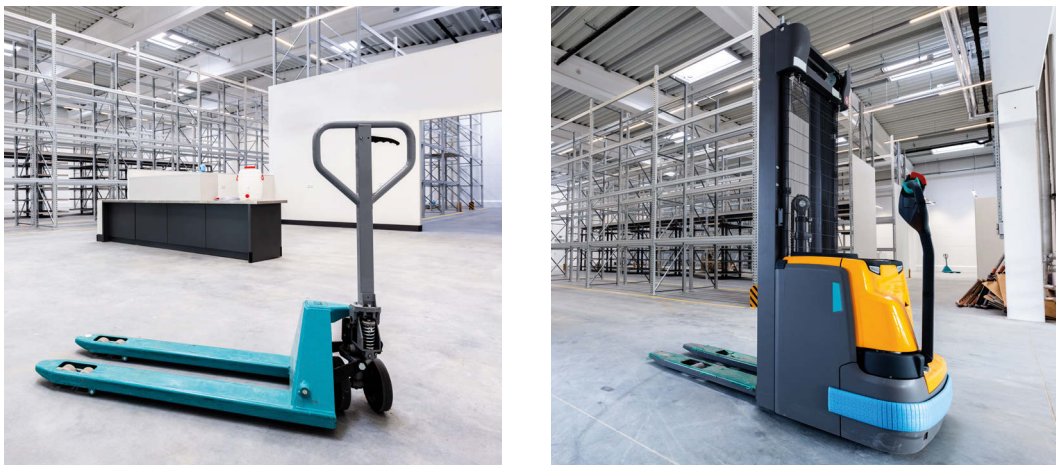
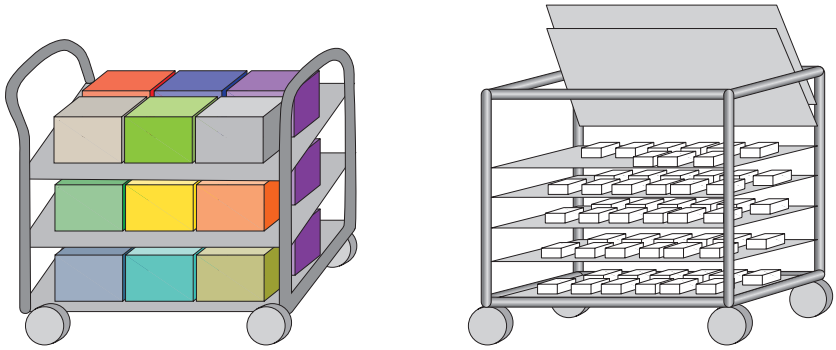


Figure 11.36 Manual and powered pallet jacks

11.5.3. **Pushcarts**

When moving parts in boxes or totes rather than pallet loads, the simple pushcart shown in Figure 11.37 is cheap, safe, can be operated by anyone on narrower aisles than forklifts. They are easiest to steer when only the front wheels swivel, but models with all four wheels swiveling can be useful in tight corners. Many baby strollers have four-wheel steering, but allow parents to lock the back wheels in line for easier steering in open spaces.

In another variation, carts have hinged shelves that rise when empty to facilitate access to the lower shelves. The hinged cart is an effective device for transportation but not for line side presentation, because it uses up too much space for a single item.



Mix-loaded manual pushcart **Cart with hinged shelves**

Figure 11.37 Mix-loaded manual pushcart

11.5.4. Tuggers and trains of tow-carts

Neither the forklift nor the individual pushcart is adequate to deliver thousands of items in box quantities at multiple locations every half hour, as is commonly needed in the automobile industry. The solution then is to hook up carts into a train and pull them around designated routes with a tugging engine whose driver also loads and unloads the boxes along the way. Tow-carts are often used for milk runs.



Mix-loaded train of tow-carts

Figure 11.38 Mix-loaded train of tow-carts

11.5.5. Automatic guided vehicles (AGVs)

Automatic guided vehicles (AGVs) have been around since the 1980s, but are still not commonplace in many factories. An AGV is a robot that moves along markers or wires in the floor, or navigates with vision, magnets, or lasers on the shop floor. As with any technology, AGVs are not a good fit for *every* manufacturing operation.

Figure 11.39 shows an AGV transporting a kit of parts from the Logistics department to the assembly line in a truck factory. This AGV uses no indoor navigation systems, such as modern simultaneous location and navigation (SLAM) technologies, but follows magnetic strips taped to the floor.



Figure 11.39 An AGV delivering a kit to an assembly line

A problem with AGVs is that the part of the transportation process that they automate is the movement itself, which, as discussed earlier, is less of a concern than preparing and loading the materials at the origin, and unloading and presenting them at destination.

In addition to carrying material to the assembly line, AGVs can also carry the workpiece (see Figure 11.40). This is especially useful for low-to-medium-volume heavy workpieces that must change height and angle to accommodate efficient assembly processes, such as the engines in the picture. Using AGVs instead of a fixed line also makes a line design flexible.

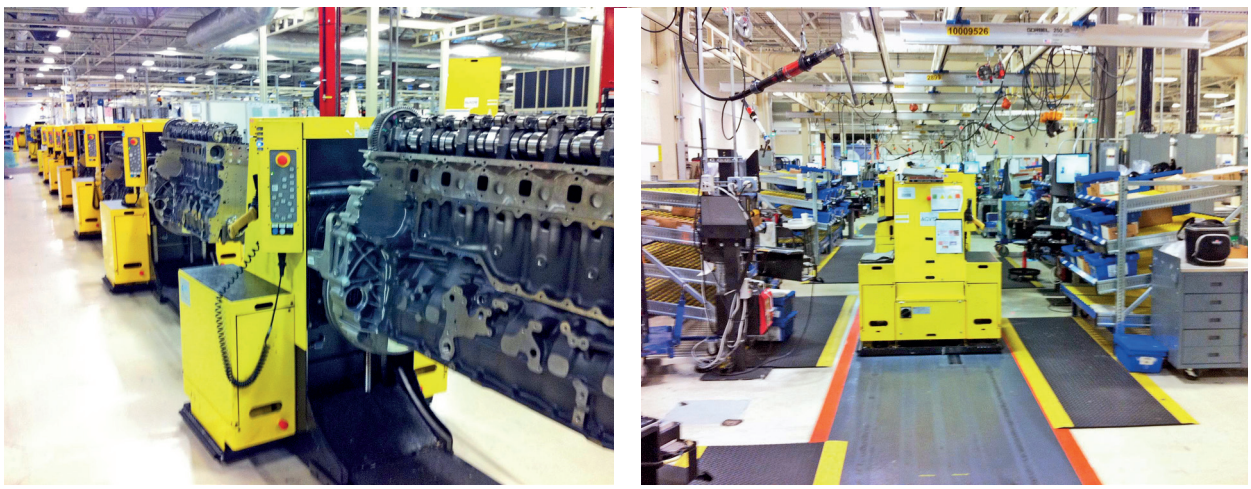
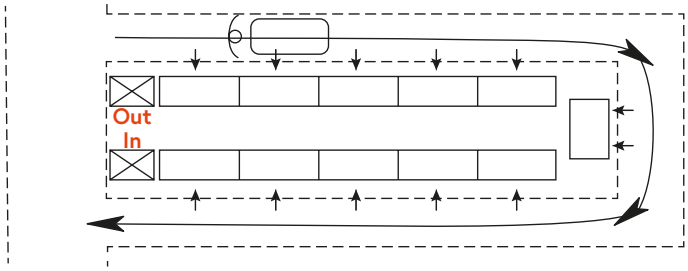


Figure 11.40 AGVs used for the final assembly of heavy vehicle engines

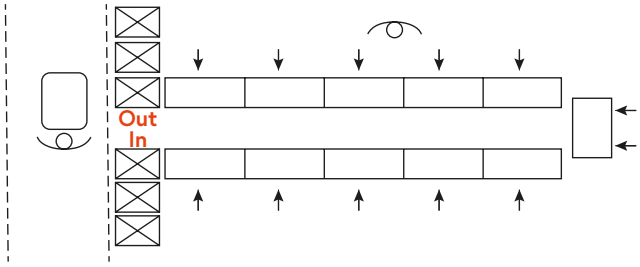
11.5.6. Part supply to assembly cells

Parts can be supplied to a cell in more than one way, as shown in Figure 11.41:

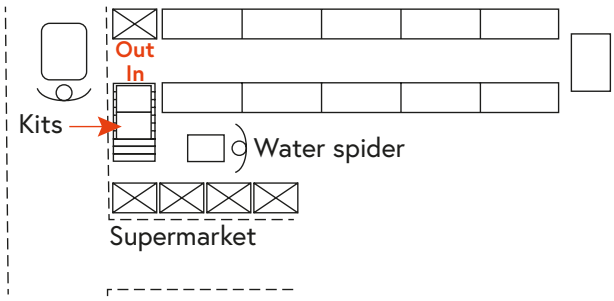
- Having the transporter deliver parts to all stations in the cell, as in Figure 11.41a, requires a transportation aisle all around, and also requires the transporters to deliver the parts in containers that are suitable for the line side.
- These constraints are relaxed in Figure 11.41b, which allows the transporter to drop, for example, pallets by the side of the transportation aisle, from which a water spider delivers bins to the stations.
- Finally, the approach shown in Figure 11.41 eliminates the need to deliver parts to any but the first station of the line, the penalty for this, of course, being double handling for all parts and the need to move complete kits through all stations.



a. Transportation aisle around the cell



b. Water spider delivers parts to all stations



c. Water spider delivers kits to the first station

Figure 11.41 Three ways of supplying parts to an assembly cell

Summary

The most essential points in this chapter were:

1. Internal logistics is comprised of all the operations needed inside a factory to deliver goods or services, except making the goods or performing the services.
2. Logistics is adding value through time, place, and presentation.
3. Internal logistics is most efficient when it flows.
4. Flow systems can be achieved by pull systems such as kanban.
5. The supply of materials to point of use should be differentiated based on the characteristics of demand and the parts.
6. Part presentation is an essential job of internal logistics.
7. Parts can be presented in kits or at lineside.
8. Conveyors are often implemented wrongly, creating monuments.
9. The choice of logistics technology must fit the task at hand.
10. Full automation of logistics is often not economical.

Test yourself

1. What are the key differences between internal and external logistics?
2. What might happen if you do not mark transportation aisles in your plant?
3. Can you describe what the following concepts have to do with manufacturing?
   - Milk run
   - Supermarket
   - Water spider
4. What is the difference between push and pull production?
5. What is a kit, and when is it better than lineside supply?
6. There are formulas for calculating the number of pull signals (for example, kanban cards), but what are their weaknesses?
7. Why is it so important to properly present parts to your operators?
8. What are the key principles of single-piece presentation and for proper kitting?
9. Can you describe five different technologies to move parts in a manufacturing plant and elaborate when they are useful?
10. Why should manufacturers seek to reduce dependency on forklifts?

Further reading

Boysen, N., Emde, S., Hoeck, M., & Kauderer, M. (2015). Part logistics in the automotive industry: Decision problems, literature review and research agenda. *European Journal of Operational Research, 242*(1), 107–120.

Hopp, W. J. & Spearman, M. L. (2004). To pull or not to pull: What is the question? *Manufacturing & Service Operations Management, 6*(2), 133–148.

Ohno, T. (2012). *Taiichi Ohno's workplace management: Special 100th birthday edition*. McGraw-Hill Education.

Roser, C. (2021). *All about pull production: Designing, implementing, and maintaining kanban, CONWIP, and other pull systems in lean production*. AllAboutLean Publishing.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

12 External Logistics

External logistics encompasses the movement of materials and products that happens *outside* the industrial estate of the company. The plant receives material from a network of suppliers and sends products to customers through a distribution network. What happens inside each of these networks affects the plant, but is often not visible to its management beyond the plant's direct suppliers and customers, usually referred to as "the first tier." Allowing each plant to know more about both its suppliers' suppliers and its customers' customers is a stated objective of supply chain management, but it is still rare.

Excellence in production and internal logistics is moot unless *external logistics* is effective at getting materials to the plant and products to market in the right condition, the right place, at the right time, and at an appropriate cost. This chapter is about the role of external logistics in *manufacturing* but not a detailed account of the operations of a logistics company, as there are other good textbooks for that. It looks at external logistics from the viewpoint of a manufacturer and covers the things a manufacturer needs to know: The function of packaging and containers, third-party logistics, milk runs, consignment stocks, and vendor managed inventory.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Understand the role of inbound and outbound logistics.
2. Design packaging and dunnage for external logistics.
3. Choose the right (returnable) containers.
4. Design systems for external logistics.
5. Decide when to use third-party logistics providers.
6. Implement vendor managed inventory and consignment stocks.
7. Understand the role of consolidation centers.

DOI:10.4324/9781351110310-16

12.1. Inbound and outbound logistics

External logistics can be split into inbound and outbound, as in Figure 12.1. Both inbound and outbound logistics are networks that consist of several tiers, or layers, of suppliers and customers. While the illustration can easily be understood, it is a gross oversimplification of real supply chains.

They are usually much longer and more fragmented, and there are usually many more parties involved and parts can go in backflows and loops. One of the most salient characteristics of manufacturing is how global it has become since the 1990s. Supply chains have become much longer and more fragmented and stretch over great geographic distances. To make this work, efficient external logistics is essential.

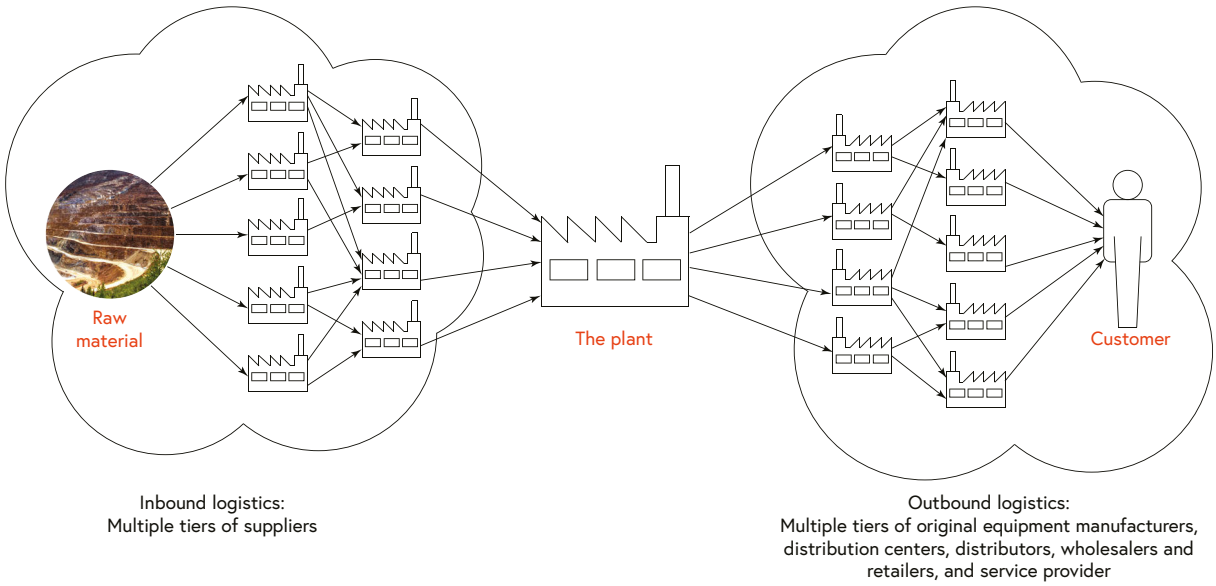


Figure 12.1 Inbound and outbound logistics

Between the nodes in the network, materials and products are physically moved (arrows in Figure 12.1). Different transportation modes can be used:

- **Pipes.** Used between buyers and sellers of liquids and gas, usually in close proximity.
- **Forklifts.** Used between buyers and sellers located in very close proximity.
- **Vans.** Used for small and rare deliveries – for example, spare parts shipped by postal services.
- **Trucks.** The dominant mode of transportation to and from a factory.
- **Rail.** Used for containers, bulk, and vehicles, but only rarely connected directly with the factory.
- **Sea freight.** Used for containers, bulk, and vehicles, but now only rarely connected directly with the factory.
- **River barges.** Used for bulk materials.
- **Air freights.** Used for quick continental and inter-continental deliveries.
- **Drones.** Used for small and time-critical deliveries between locations not linked with efficient transportation routes – for example, blood samples between hospitals.

12.2. **Multimodal and intermodal transportation**

When deliveries change transportation mode while underway, it is called *multimodal or intermodal transportation*. Multimodalism simply means transport by multiple modes, and has existed for hundreds of years since early cargo sea shipment and the first years of the railway. Intermodalism, on the other hand, has grown since the 1980s, and involves standardised and integrated interfaces between modes. In other words: Intermodal transport is similar to multimodal transport but puts more emphasis on connectivity of different transport modes. OECD defines intermodal transportation as follows:[1]

> **Definition**:
> Intermodal transport indicates the use of at least two different modes of transport in an integrated manner in a door-to-door transportation chain.

For intermodal transportation networks to be efficient, the main goods flows are routed through centralised nodes where the goods are efficiently transhipped to other carrying units. The transhipment can be between similar modes (rail–rail, road–road, etc.) or different modes of transport (rail–road, sea–road, etc.). Thus, the intermodal *terminal* is the key to achieve efficiency in intermodal networks. It is at the terminal that the "inter"-aspect of intermodality is realized (see Figure 12.2). Intermodalism as it exists today is far from seamless.

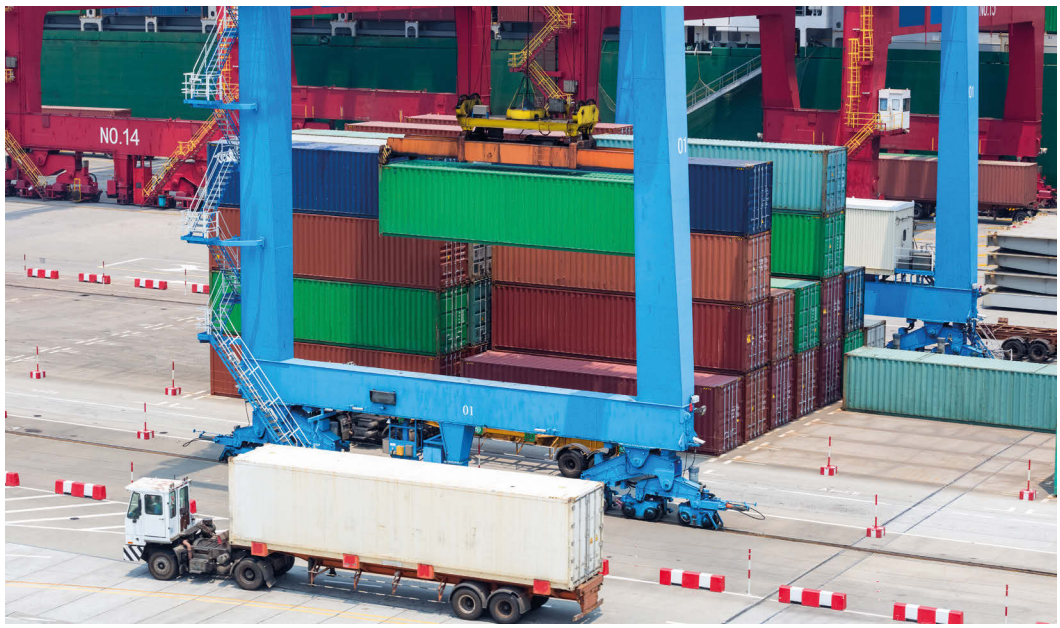


Figure 12.2 Loading of container from ship to terminal to truck

Usually what transportation mode to use is not the concern of the manufacturer, but the job of the logistics provider. For outbound logistics, the manufacturer is concerned about the price, quality, and time of delivering a product to its customers. For inbound logistics, the manufacturer is mostly concerned with having the part or material at the right time in the right quality at the right place. Yet, they need to know how the goods will be transported to understand packaging requirements.

1 OECD. (2002). *Benchmarking intermodal freight transport*. Paris, Report number: E112021.

12.3. **Containerization**

Today, almost all products are not "made in Germany" or "made in China," but "made in the world."[2] The long global supply chains are enabled by global logistics. Logistics is a truly global activity. Today, all locations where people live can be reached through postal and logistics services – and it has been so for hundreds of years. What has changed is the speed and cost by which logistic services can be carried out. In 2020 and 2021, however, the COVID-19 pandemic exposed vulnerabilities in global logistics.

While it took months to ship things across continents before the Industrial Revolution, today it can be done in weeks per sea freight and hours per air freight. More important is the plummeting cost of international transportation. Shipping rates vary widely depending on location, destination, and shipping company, but, to give a rough estimate, in 2021, a 20-foot container shipped from the US to China port-to-port could perhaps cost $500 to $5000 depending on bargaining power. A large manufacturer of industrial goods will negotiate low rates. Consider that a 20-foot container can, theoretically, hold 100 washing machines or 400 flat-screen TVs. The shipping costs then become $5 per washing machine and a ridiculously low $1.25 per TV. Of course, these are only simple estimates, and, in addition, administration, tariffs, and transportation to and from the ports will be added. It illustrates, however, the motivation of many manufacturers to move production to low-cost countries.

One particular technological development has played an important role: *The container*. The efficient changeover between transportation modes was enabled by the invention of the shipping container in the 1950s (see Box 12.1). Figure 12.3 shows how ships were loaded and unloaded prior to Malcolm McLean's invention:

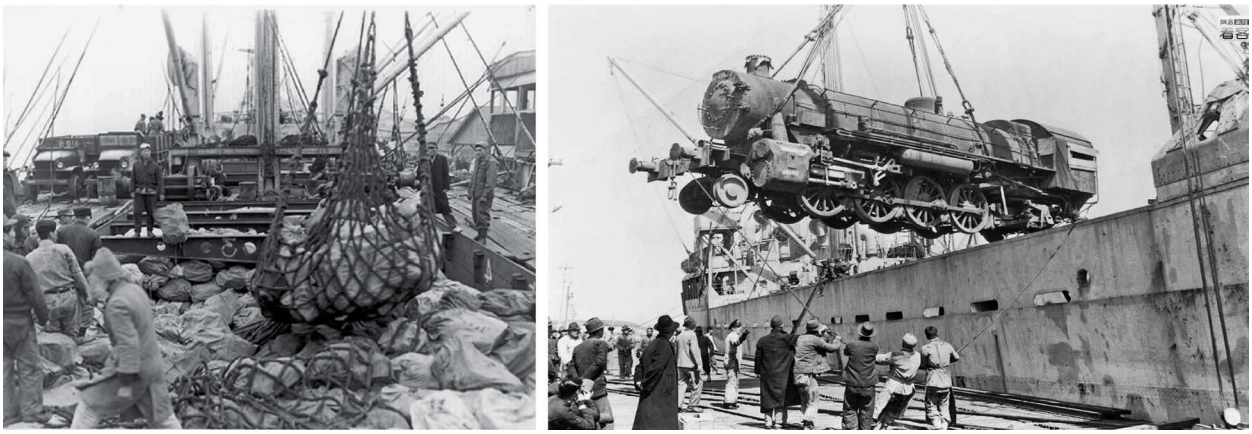


Figure 12.3 Ship unloading before containers

2 Ferdows, K. (1997). Made in the world: The global spread of production. *Production and Operations Management, 6*(2), 102–109.

Box 12.1 **The container**

Around 95% of global trade is now transported in containers. The standardized container has minimized transport handling costs and been an important driver for global trade. The container was patented by Malcom McLean in 1956. McLean had observed long waiting times in seaports. For example, in the 1930s, armies of dock workers were busy unloading the truck (or ship), repacking, storing goods, and loading the ship (or truck). To unload a 5,000-ton ship required 60 men for a week. The introduction of the standardized container and container handling equipment reduced the time needed for intermodal shifts to a small fraction of what it used to be.



Figure 12.4 Global logistics and intermodal transportation use containers (Credit: David Dilbert/pexels)

Figure 12.4 shows the standard container sizes and Table 12.1 gives standard measures for the two most typical containers. In 1968, a common ISO standard set the dimensions for a standard container: The twenty-foot equivalent unit (TEU). TEU is still used today as a unit for ship capacities and port handling volumes. Extra tall shipping containers called high-cube containers are 9.5 feet (2.89 meters) high. To give an impression of the size, a 20-foot container can fit a car and boxes from a garage, or all the furniture and personal goods of an 80-square-meter apartment.

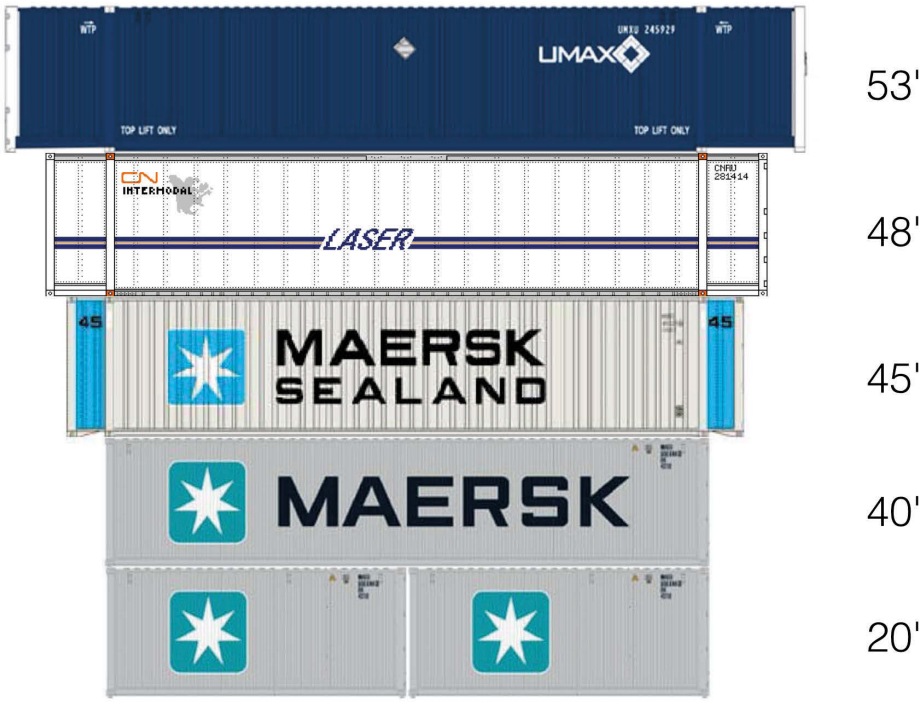


Figure 12.5 Container sizes (Credit: Kcida10, CC BY-SA 3.0)

Table 12.1 Standard measures of 1 TEU and 2 TEUs containers

Measures	1 TEU container	2 TEUs container
Size (LxWxH)	20ft x 8ft x 8.5ft 6.06m x 2.43m x 2.59m	40ft x 8ft x 8.5ft 12.2m x 2.43m x 2.59m
Total internal volume	1,172 ft³ 33.2 m³	2,389 ft³ 67.6 m³
Pallet places	10 GMA pallets in 1 tier 11 Europallets in 1 tier	20 GMA pallets in 1 tier 23-24 Europallets in 1 tier
Weight of empty container	2,300 kg	3,750 kg
Net shipping load	28.3 tons	26.9 tons

12.4. **Incoterms**

Incoterms are a set of rules that define the responsibilities of sellers and buyers for the delivery of goods under sales contracts. "Incoterms" is an acronym standing for international commercial terms. Incoterms are governed by the International Chamber of Commerce (ICC). Incoterms are widely used in international trade. It is of course extremely important to know your rights and responsibilities when receiving or shipping goods. Below is a summary of the most common codes in the 2020 Incoterms that replaced the 2010 incoterms. For all the details check the webpage of ICC: *https://2go.iccwbo.org*.

Group 1. Incoterms apply to any mode of transportation:

- **EXW (Ex Works)**. The Seller allows the Buyer to collect the goods at an agreed point, usually the seller's factory or warehouse. The Seller's only responsibility is to prepare the goods for pickup by the Buyer's carrier. The Buyer holds all the risk of shipment, hence EXW is usually not preferred by buyers.
- **FCA (Free Carrier)**. The buyer and seller define a handover location. If the parties name the seller's facility, the seller is responsible for loading the goods onto the initial carrier and the buyer is responsible for transport, freight charges, and risk of loss thereafter as well as export and import responsibilities.
- **CPT (Carriage Paid To)**. The buyer and seller name a location on the buyer's side to which the freight is prepaid by the seller. The import responsibility is the buyer's obligation, while the seller is responsible for transport, freight charges, risk of loss, and export obligations.
- **CIP (Carriage and Insurance Paid To)**. The parties name the place on the buyer's side to which the freight is prepaid by the seller. Seller carries all of the responsibility, including providing insurance against loss or damage, and import responsibility and unloading from the final carrier are the buyer's only obligation.
- **DAT (Delivered At Terminal)**. The buyer and seller identify a terminal on the buyer's side to which the goods are delivered. Seller carries all of the responsibility, except import responsibility falls on the buyer as well as unloading at the terminal.
- **DAP (Delivered At Place)**. The parties name a stipulated destination on the buyer's side where the goods are to be delivered. Seller carries all of the responsibility, except import responsibility falls on the buyer as well as the obligation to unload once it reaches the designated destination.
- **DDP (Delivered Duty Paid)**. The parties name a Buyer's location to which the goods are delivered. All responsibility falls on the seller, except for obligations of unloading once the goods reach the Buyer's location.

Group 2. Incoterms apply only to sea and inland waterway transport, and are followed by a named port of shipment, as in "FAS Rotterdam":

- **FAS (Free Alongside Ship)**. The seller is responsible for the cost of transporting and delivering goods to the export point, but the responsibility shifts to the buyer once the goods are unloaded at the export point.
- **FOB (Free On Board)**. The seller has responsibility until the goods are loaded on the transport vessel, then all responsibility shifts to the buyer.
- **CFR (Cost and Freight)**. Buyer assumes responsibility for loading the goods on the truck at the place of import, transport to the destination, and import responsibility.
- **CIF (Cost, Insurance, and Freight)**. Same as CFR plus an additional requirement that the seller purchases insurance against the risk of loss or damage to goods.

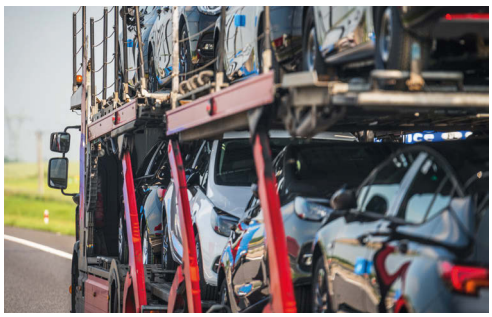
For container shipment, it is also usual to separate Less than Container Load (LCL) from Full Container Load (FCL):

- LCL refers to shipments that require less than a 20-foot or 40-foot full container.
- FCL refers to shipments where you pay for an entire shipping container. Shipping companies generally prefer when customers pay for FCL as it simplifies the logistics processes. This means no one else's goods ship in your container.

12.5. Packaging for external logistics

Except for ships, airplanes, buses, and trucks that are themselves means of transportation, few products travel unpacked. As seen in Figure 12.6, the packaging used for different ranges from nothing to multiple layers that occupy a much greater volume than the product itself:

- Cars travel "naked" – or when completely new, partly covered in plastic covers – on trucks custom-designed for this purpose. This is due to their size, their ability to move on their own power, the fact that they are intended to be exposed to the elements, and the volume of transportation they generate.
- Motorcycles and scooters travel in crates that can go inside any standard truck. The crates require labor to assemble, load, unload, and recirculate. The much lower volume than cars does not justify custom trucks.
- Seeds travel in big bags on standard pallets handled with standard logistics equipment.
- Microprocessors travel in individual boxes that are many times their size, in multiple layers of protection against electrical as well as mechanical shocks, and documentation. Not only does packing and unpacking use labor, but the boxes are disposable.



a) Car transport



b) Motorcycles



c) Seed transport



d) Intel CPU in original packaging (Credit: CC BY-SA 3.0 by Dsimic)

Figure 12.6 Packaging for a variety of industrial products

12.5.1. Need for packaging

Since packaging does not change a manufactured product in any way, it is tempting to dismiss it as a "non-value-added" activity, but it does not let itself be dismissed that easily. There are many reasons why it is necessary, which are not about to go away. The following is a partial list:

1. Protecting the product against hazards:
    a. **Mechanical damage.** This is the most obvious. During loading, transportation, unloading, storage, retrieval, etc. the product is at risk of being dropped, hit, or shaken.
    b. **Electrostatic discharge (ESD).** This mostly affects integrated circuits and printed circuit boards handled inside a plant. You prevent it by using metalized bags or conductive totes that enclose the part in a Faraday cage. Although one might expect electronic components to also require protection against magnetic fields, it is in practice much less of a concern than ESD.
    c. **Corrosion.** Automotive parts with exposed metallic surfaces are often packed in grease for protection against corrosion, particularly when exposed to sea air during maritime transportation.
    d. **Contamination.** Ingredients going into pharmaceutical or cosmetics products require sterile packaging that seals them off from biological contamination.
    e. **Dust.** Parts and products have varying levels of cleanliness requirements, the most stringent being those of semiconductor wafers, which must not only be fabricated in cleanrooms but also moved in containers that keep dust out.
    f. **Heat and light.** Products that are sensitive to heat and light include not only the obvious foodstuffs and photographic materials but also painted surfaces that can crack or be discolored.
    g. **Theft.** Products that are both expensive and small are often protected from theft by outsize packages that make them more difficult to conceal.
2. Holding the product(s) in place or shape.
3. Protecting the environment against hazardous materials, particularly in the form of leaks of chemicals and radiation from nuclear materials.
4. Protecting handlers against sharp edges the product may have. Knives have sharp edges, but so do machined workpieces prior to deburring.
5. Assuring that products are new and untampered.
6. Gift wrapping.
7. Presenting industrial goods for convenient picking.
8. Communicating information about the product through labels, kanbans, color-codes, or by making the content visible by transparency as in bottles or wire-mesh baskets.
9. Placing a cap on inventory. If there is a fixed number of packages in circulation, then the amount of inventory of the item in those packages is limited to their content.
10. Marketing and sales. Changing the wrapping of an old product can make it appear new and increase its sales. A nice packaging design can enhance the products' appeal on store shelves.

Stop and Think 12.1
The above list gives good reasons to take packaging seriously. What do you think are the main downsides or challenges of packaging?

12.5.2. Choosing between disposable and returnable containers

While disposable containers will continue to be the norm for consumer goods, their use in manufacturing logistics is being challenged by returnable containers. If we consider the case of cartons used in shipping parts from a component supplier to an assembly customer, it involves the following activities that would not be needed with returnable containers:

1. The supplier carries out a commercial transaction for each purchase of cartons.
2. The cartons come themselves in palletized bundles that need to be unpacked prior to being brought to their point of use.
3. Operators in the supplier plant need to erect the cartons and insert dunnage.
4. In the customer plant, used empty cartons need to be collapsed and bundled, sometimes requiring the kind of heavy equipment shown in Figure 12.7.
5. The customer bundles empty cartons and ships them either to landfills or to recycling plants, to be turned into new cardboard, building insulation, magazine paper, or newsprint.



Figure 12.7 Processing of empty cartons in a cardboard bailer (Credit: HSM)

In addition, disposable containers place no limit on the amount of inventory in circulation between the supplier and the customer. From a production control standpoint, cartons can be treated as a resource in unlimited supply.

Returnable containers must also be bought, but purchases are two orders of magnitude less frequent than for disposable containers. Contrary to disposables, their supply is perceptibly finite: Management decides how many containers are in circulation, and, in principle, the quantity of any given item in circulation between supplier and customer cannot exceed what fits in these containers. When daily volumes fluctuate within a narrow band around a stable average, this is a means of preventing overproduction. It is, however, not an advantage when demand suddenly doubles or is cut in half, or when customers require suppliers to ship in disposable containers when returnables are short.

The process of returning empties to where they can be refilled involves the following:

1. **Sorting.** The customer organization needs to sort the empty containers by item and ship them to a Container Management Center (CMC) to route them accurately back to suppliers. Suppliers also track containers by customer. To production, empty containers on the shop floor are a nuisance to be disposed of as quickly as possible, and many organizations respond by just throwing them together, relying on the logistics organization, perhaps in a CMC, to separate them by supplier and item. The logic behind this policy is that, if third-party materials handlers off-site are paid 1/3 the wages of the customer's employees doing similar work, then it makes sense to offload this task. The reality, however, is that this sorting is unnecessary if the customer organization doesn't scramble the containers in the first place.
2. **Storing.** Storage space is needed by both customers and suppliers for empty containers before shipping and after receiving. In some cases, in urban areas, open storage sheds are either not secure enough or not allowed by local zoning ordinances, in which case warehouse space inside the plant is set aside for empty returnable containers. Note that the use of supplier milk runs reduces the storage requirements for empty returnables by one order of magnitude compared to the single-item full truckload approach.
3. **Transporting between plants.** Empty returnables must be trucked from the customer plant back to the supplier. Here again, supplier milk runs are key, because they allow the containers to hitch a ride in trucks that would otherwise travel the same route empty.
4. **Moving within the plant.** Empty returnables must be delivered to the filling point, and in-plant milk runs also provide the opportunity for empty containers to ride on vehicles that would otherwise travel empty.

> **Stop and Think 12.2**
> Do you use returnable or disposable carriers to get your groceries home from the supermarket? Why?

12.5.3. Economics of returnable and disposable containers

In their promotion of the concept, the suppliers of *returnable containers* emphasize the following:

1. **Packaging quality.** Plastic containers, particularly stackables with part-specific dunnage, are simply better protection against handling damage than disposable, corrugated cardboard boxes. The best way to quantify this is to measure the results of a pilot implementation and extrapolating to the entire product mix.
2. **Costs.** While a switch to returnable containers eliminates some activities, it requires others, and we have no basis to say that the cost of handling returnables is always lower than that of disposables. These operational costs, however, can be small compared to the costs of the containers themselves, and this is where the differences appear, as discussed in detail below.
3. **Environmental responsibility.** Disposable containers generate waste that returnable containers don't. Some managers view the switch to returnable containers as a way of "greening" the company. The public relations benefits of this may be substantial but are difficult to quantify.

If a disposable package costs $2 per unit and a returnable $40 per unit, then the returnables become cheaper once they have been used more than 20 times, assuming handling costs are similar. The number of times a returnable container can be used is clearly key to its economic value, but its frequency of use is equally important. The payback period on our $40 returnable container is 20 years if used once a year,

but 4 weeks if used daily. Household coffee makers are sold to consumers in disposable packages because they don't buy one every day. A car assembly line, on the other hand, uses 1,000 sets of rearview mirrors every day or 50 returnable containers with 20 sets each.

Customers do not like the idea of paying to ship empty containers back to suppliers. This problem can be solved with supplier milk runs. The conditions for making returnable containers pay off are therefore as follows:

- Use them only for items with steady demand.
- Handle them with care to maximize the number of uses.
- Keep their numbers low enough for them to rotate frequently between customers and suppliers.
- Take advantage of available transportation. For example, take back empty containers when delivering new products.

12.5.4. **Types of returnable containers**

Returnable containers come in many shapes and sizes, to fill various needs. Three typical types are:

1. Stackable containers
2. Nestable containers
3. Collapsible containers

Stackable containers

Stackable containers, as shown in Figure 12.8, are the most commonly used in local supplier milk runs. They take up as much space empty as full, but that is not a problem in the absence of return freight opportunities from the customer back to the supplier. The containers' vertical walls enable them to hold multiple identical layers of parts, and the corresponding dunnage can remain in the empty returning container. In addition, stackable containers have lips intended to hold the containers above securely and do not need lids.



Figure 12.8 Stackable containers

Nestable containers

By contrast, the empty nestable containers shown in Figure 12.9 occupy only a fraction of the volume they take up when full, but they cannot hold multiple identical layers of parts and the dunnage must either be disposed of or transported back to the supplier separately, which can nullify the space savings. Their stacking stability is also inferior to that of stackables and they require lids. The main limitation of nestable containers is their tendency to ride up when bumping into each other on a conveyor. However, dunnage and stability are more important issues.



Figure 12.9 Nestable container

Collapsible containers

Pallet-size wire baskets are collapsible containers and used in many machine shops, but are often too large for the size of parts they contain. Smaller, collapsible containers, such as shown in Figure 12.10, are used, for example, for transporting automobiles parts across oceans.



Figure 12.10 Collapsible container

By collapsing to a fraction of their erected size, these totes allow their users to fill ocean-going containers with return freight, and they commonly establish trading companies for the purpose of finding such freight. Collapsible containers are sometimes also used domestically, but they have drawbacks:

1. They obviously do not allow dunnage to recirculate with them.
2. Operators describe them as "finger pinchers." Erecting and collapsing these containers quickly and safely requires special skills and dexterity.
3. Due to their design, they are less robust and stable.

12.5.5. Scope of use of returnable containers

The containers used to transport parts between facilities exist for the purpose of handling and protecting parts in transit, not to present parts appropriately to production operators at the customer site. On occasion, they may be usable for this purpose but that is not common. There usually is either a size mismatch with the need of the line or a part presentation issue; both problems are discussed in the following sections.

The first and most common problem is that transportation containers are *too large for lineside use*. One full pallet-size bin, as unloaded from the truck, may contain two weeks' worth of a part that goes into one out of ten products assembled on a given line. While other products are being assembled, their parts need the line side space taken up by the bin, which must, therefore, be removed, returned to stores as a "partial," to be retrieved when needed later. Since parts are often just heaped into these bins, knowing how many are in a partial bin is practically impossible. Therefore, in addition to creating extra handling work, the manipulations of partial bins make inventory accuracy unmaintainable.

When there is no alternative to receiving parts in large bins, in-plant logistics needs to break them down for lineside delivery into bins that are small enough to be allowed to stay in place until empty, so that the parts' trip from the warehouse to the line is strictly one-way. The content of the warehouse is decremented by a whole number of small bins and never has to be adjusted back up for partials afterward. As shown in an example in Figure 12.11, this also results in saving about 60% of the floor space around the cell, using narrower aisles, and eliminating forklift traffic from the production area. The situation shown in Figure 12.11a commonly emerges in the early stages of conversion to a lean production system as a result of cell implementation. Subsequent logistics improvements then allow the plant to migrate to Figure 12.11b.

While breaking down large bins add handling labor upfront, it is justified by the following benefits:

- Reduction in the amount of labor required later to handle partial bins.
- Simplification in information system requirements, by the elimination of return transactions for partials.
- Enhanced shop floor visibility. The flows of materials are easier to follow when they go in only one direction.
- Improved inventory accuracy. The actual number of parts in partial bins is often inaccurately recorded.

The second problem is *part presentation*. Parts should be presented to assemblers unpacked, within arm's reach, with their smallest dimensions facing out, and oriented for easy installation, with single-piece presentation and kitting as needed. Transportation containers are designed to meet other needs, and, as a consequence, are frequently inadequate by the line side.

Some plants transfer parts to lineside bins right at Receiving and the transportation containers never make it to the inside. This is mostly done with disposable, cardboard containers, and where the dust associated with cardboard on the shop floor would hurt the products. This strategy is used, for example, by semiconductor equipment maker Applied Materials. Other plants deliver transportation containers to supermarkets near the production lines, where water spiders unpack parts and kit them for delivery to the line.

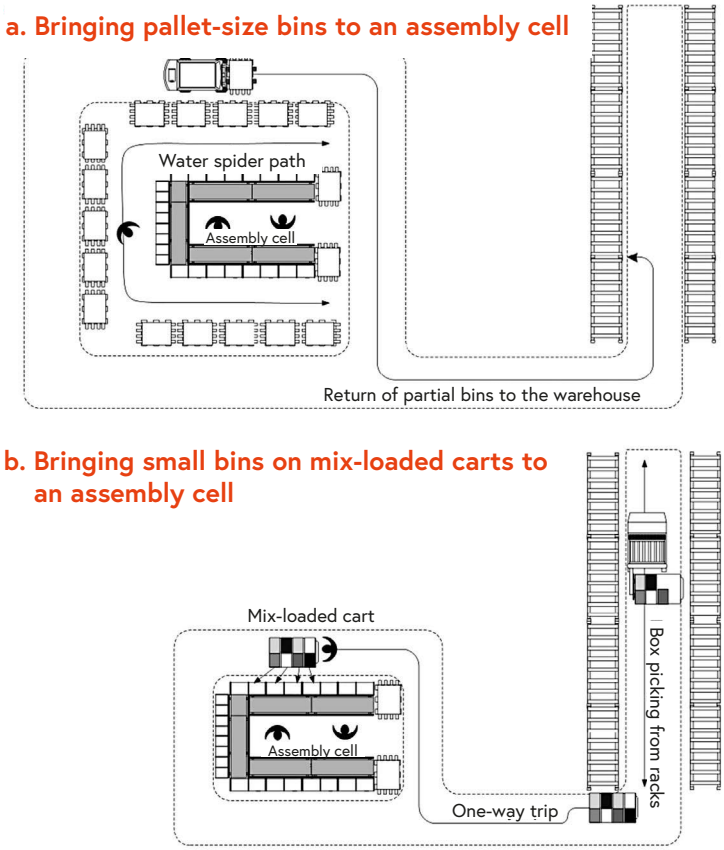


Figure 12.11 Impact of bin size on delivery to an assembly cell

12.5.6. Sizing the containers

Converting all handling from large to small bins increases the packaging cost per unit, but it only results in a cost increase if the small bins are used to handle the *same quantities* of parts. If, as shown in Figure 12.12, the small bins are used to reduce the number of parts handled, their use does not result in an increase in overall packaging cost. The point of reducing bin size is not to replace one large bin with 27 small ones but with 1 small one. It is difficult to reduce the quantities handled when using large bins. Small bins are necessary to make that possible.

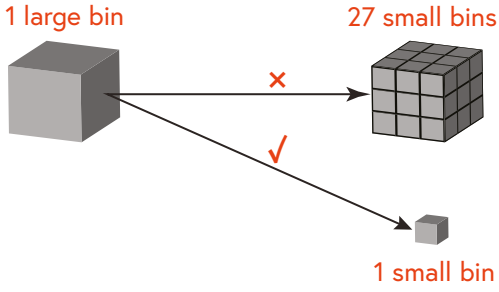


Figure 12.12 Large versus small bins

12.5.7. **Types of dunnage**

Figure 12.13 shows several dunnage options, in increasing order of cost and function. None of these options should be excluded. The key point is that this level of operational detail should be designed based on an analysis of the needs.



a. No dunnage b. Disposable dividers c. Reusable dividers d. Part-specific dunnage

Figure 12.13 A few dunnage options

Provided they are not subject to damage as a result, parts can simply be heaped in a returnable container without any dunnage, as in Figure 12.13a. This has the following drawbacks:

- **No mistake-proofing.** Wrong parts can easily contaminate the bin.
- **Counting by weight only.** Weighing the bin is the only practical way of knowing how many parts are in it. It requires a scale and may not be sufficiently accurate and precise.
- **Random orientation of parts.** Randomly oriented parts take longer to pick by hand and are a challenge for pick-and-place robots.

Disposable cardboard dividers, such as shown in Figure 12.13b, provide more structure by, for example, separating the parts into sets that go into one unit of finished goods. They also make parts easier to count, but they provide neither mistake-proofing nor orientation control.

The reusable dividers in Figure 12.13c go one step further. They provide limited orientation control, but it is still possible to insert a part upside down and to insert the wrong part. Only the part-specific dunnage in Figure 12.13d provides easy counting, mistake-proofing, and orientation control. The downside is the time spent placing the product into the dunnage.

> **Stop and Think 12.3**
> Consider the four types of dunnage in Figure 12.13, which one would you prefer for the following parts:
> a) Diamonds for jewelry selling 100 pieces a month
> b) Feathers for shower curtain poles selling 1 million pieces a year
> c) Aluminium back covers for smart phones selling 1 million pieces a year
> d) Special graphic cards (GPU) for computers selling 1000 pieces a month

12.5.8. Preparation of shipping-ready containers

A manufacturer can often not control the type of containers customers want parts shipped in, and even plants that make every effort to use returnable containers both internally and with suppliers often must package their finished goods in disposable containers if the customers require it.

On the shop floor, the flow of disposable containers needs to be organized as a production process and synchronized with the production of the goods. Strictly speaking, packaging goods is part of the production process and not logistics, but, in most plants, at least part of the responsibility for it is in the Logistics organization. Sometimes the work is divided between Production and Logistics. For example, packing the detergent powder into boxes and grouping the boxes into display cases are Production tasks, but Logistics palletizes the cases and shrink wraps the pallets.

There may be a variety of reasons why packaging is a split activity between Production and Logistics:

1. The pallet configuration for shipment is not known at the time of production.
2. The product is customized at the last minute. For example, individually boxed items sent to department stores require store-specific price tags.
3. Production's outgoing quality may not be trusted, and Quality Assurance demands a final inspection prior to shipment.

The automatic packaging machines used with mass-market consumer goods like detergents or personal products pick, erect, and fill packages one at a time. Manual packaging (Figure 12.14), on the other hand, is often not designed with the same care as other production operations and involves operators building a stock of erected boxes ahead of need. This practice is unsound for the following reasons:

- It adds storage and retrieval work between erecting and filling a box. Even though it is done in the name of "efficiency" is adds labor to the process.
- Boxes take up about 20 times more space erect than flat. Empty erect boxes not only take up valuable line side space but also clutter the workspace and reduce visibility.
- The risk of closing and shipping an empty or partially filled box is higher when there are many empty erect boxes around than when the only erect box around is the one currently being filled.



Figure 12.14 Manual erection of cardboard boxes in a confectionary factory

12.5.9. **Ownership of containers**

When container ownership is settled, it is eventually reflected in the price paid by the customer. If customers buy the containers, they pay correspondingly less for the goods than if they ask the suppliers to make that investment. The issue is not which side profits the most but what incentives each side has to be diligent and effective in managing the flow of containers.

When returnable containers were used to deliver milk or beer to consumers, they were owned by the suppliers, who had to motivate the consumers to return them, for example by giving them a few cents back for each empty bottle. By contrast, returnable containers used between manufacturing companies are often owned by the customer side. Some customers ask suppliers to design and buy the containers and then buy them back from the supplier over time.

Among the reasons for this, the most fundamental is that it ensures that the containers do recirculate and are handled with care. Customers are neither careful nor prompt in returning containers owned by their suppliers. Suppliers, on the other hand, are keen to "take care of their customers" by handling containers as the customers wish and are in a hurry to return them full of their own products. In addition, customers often specify protection requirements for the parts they buy, and these requirements are easiest to meet if the customer owns the containers and the dunnage.

On the other hand, if any side owns all the containers used for a given product, then the other side has no strong incentive to reduce the number in circulation. Ownership of containers can also be split. Assume, as shown in Figure 12.15, that there are 28 containers in circulation, with on average 10 in the customer plant, 8 in transit, and 10 in the supplier plant. Then the customer may own 18 of the containers and the supplier the remaining 10, which ensures that both sides are motivated to make do with as few containers as possible.

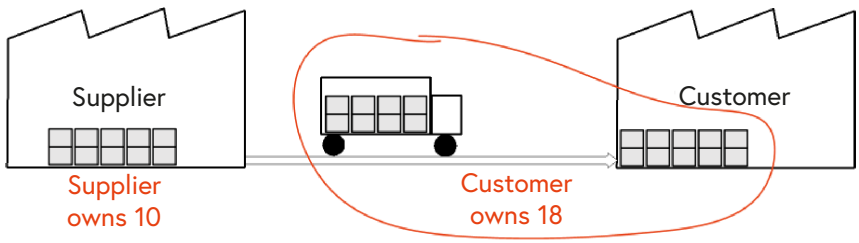


Figure 12.15 Ownership split for returnable containers

The supplier cleans, maintains, repairs, and replaces the containers as needed. Supporting suppliers in this job is unfortunately not a widespread tradition, which is reflected in the condition of empty containers arriving at supplier plants. It is not uncommon for them to be smeared with oil, broken, or commingled with other suppliers' containers. Customers also tend to use the suppliers' yards as free storage of extra containers. Most of the costs incurred by suppliers for these reasons would be avoided if the customers paid more attention to the flow of returning containers. They don't, because they are busy attending to their own customers and feel that their suppliers should do the same. This attitude towards suppliers, however, adds the cost of these operations to the price of the products the customer pays. Compared with this approach, working together with the supplier to avoid generating extra work is the way for the customer to get lower prices while allowing the supplier to make a profit.

12.5.10. Countermeasures against container shortages

Having a fixed number of containers in circulation places a cap on the inventory of the corresponding item. There cannot be more material in the loop than can fit in the containers available, and, if container circulation is properly managed, there should be no situation where the supplier needs to ship parts but has no container with which to do it.

In the Toyota supply chain, this is indeed a rare event. When it does occur, the supplier's response is to notify the customer plant and request an emergency shipment of empty containers. The supplier is neither expected nor allowed to revert to disposable containers. American car manufacturers do not have this policy. Figure 12.16 shows a finished goods flow rack with several rows of disposable containers behind the returnable containers that are standard for this product.



Figure 12.16 Disposable containers used in conjunction with returnables

If the elimination of handling damage is a stated goal of the move to returnable containers, then how can the customer allow some of the goods to be knowingly shipped in containers that offer inferior protection? Concurrently using both types of containers for the same item also complicates the task of measuring the reduction in handling damage, and requiring suppliers to maintain stocks of disposable containers drives costs up.

12.6. Container tracking

When hundreds or thousands of items are handled in returnable containers, keeping track of how many of which type are located where is a considerable task. There are two complementary approaches:

1. Visible management through color codes, labels, and signs
2. Computer-based information system for track and trace

12.6.1. **Visible management of containers**

Color coding is a time-honored way of enhancing the distinctiveness of otherwise similar-looking objects, thereby making them less likely for handlers to confuse. It is easier to use with returnable containers, which come in multiple colors unlike cartons that are brown unless custom-printed. There are, however, limitations to keep in mind:

1. Boxes come in a limited number of colors. In addition, the human eye can tell apart only so many colors reliably over time, as they fade and boxes get dirty. As a result, colors have to be assigned to categories of parts rather than individual items.
2. There are cultural constraints on color choices. Red, for example, is commonly used for defectives, and should therefore not be used for anything else.
3. Color-coding limits flexibility in the use of containers. Regardless of changes in volume, green containers cannot be reconfigured to hold the same parts as blue containers.

These limitations are sufficiently severe for some manufacturers to decide that color-coding is not worth the trouble. Labeling requirements are issued by customers. Ford, for example, requires hot stamps on two sides of each container, bearing the following:

- The supplier company's name, city, and state
- A Ford container code
- A "ship from" location code

The information actually carried on container labels, however, varies wildly. Figure 12.17 shows empty nestable containers delivered to a supplier that belong to other suppliers and are labeled inconsistently.

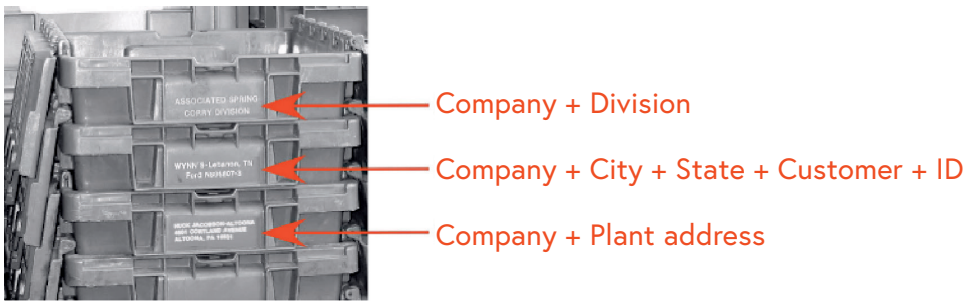


Figure 12.17 Returnable containers from multiple suppliers

In addition to this permanent identification, containers also bear stick-on labels designating the parts they carry, and, if the kanban system is used, a plastic pocket or other devices to hold a kanban. Figure 12.18 shows an example of an in-house *production kanban* held in a container's card pocket in an injection molding shop. In this example, the card pocket is used to hold other data as well. The backflush card goes to Accounting and the traveler to Quality Assurance.

It is easy to attach such a card pocket to plastic totes as in Figure 12.18, but other types of containers, such as wire baskets, are more of a challenge. Finding a way to attach kanbans to containers that is both convenient and sturdy is often a challenge when implementing the card-based kanban system. Alternatives to card pockets include velcro, magnets, and the kind of clips that are used to attach badges to clothing.

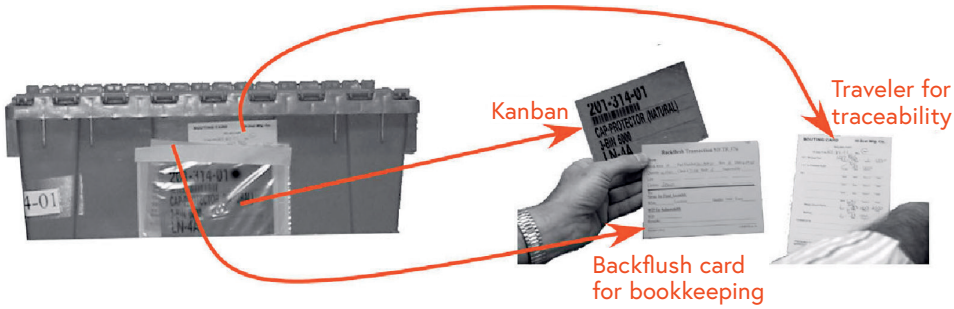


Figure 12.18 Returnable container with kanban in the card pocket

Signs on the walls and markings of storage locations assist in moving containers appropriately. As everywhere else, the actual visibility of the signage and the maintenance of its accuracy require sustained attention. If just a few of the signs on the shop floor are inaccurate, the credibility of all signs is destroyed in the minds of materials handlers and operators.

Visibility tools, however, are insufficient and must be supplemented by an information system. For some purposes, we don't need to know the container status in more detail than shown in Figure 12.19, and in terms of numbers of containers rather than the fate of individual containers.

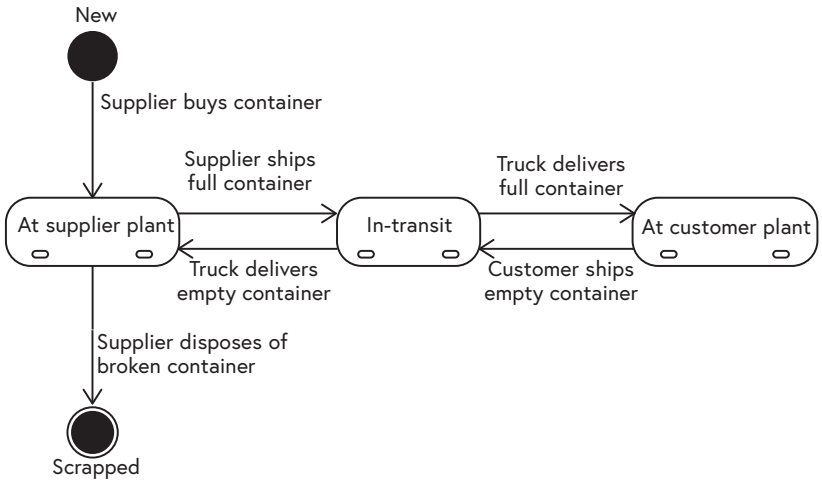


Figure 12.19 Container states, high-level view

The number of containers in each state can be gleaned from data already collected through basic production control transactions. In routine operations, the numbers of containers at the customers, in transit, or at the suppliers should vary within bounds around an average, and large deviations can be used as early warnings of shortages. Also, the average number of times containers are reused before being scrapped can be inferred from this data. In steady-state, we have:

$$Container\ life\ expectancy = \frac{Number\ of\ containers\ in\ circulation}{Scrap\ rate}$$

If the supplier scraps 5 containers every month out of 50 in circulation, then, on the average, a container lasts 10 months. If the supplier ships 500 boxes per month, then each box makes 10 round trips per month or 100 round trips in 10 months. We know then that a box is reused 100 times on the average.

12.6.2. **Semi-automatic and automatic track and trace**

Track-and-trace is the process of identifying past and current locations of items. "Track" is identifying where an item is located just now; "trace," the ability to identify where it has been. If tracking is done well and the data is stored appropriately, traceability follows.

Tracking individual containers is usually not worth the trouble if it requires manual data entry. However, items become easily tracked if fitted with an auto-ID technology such as barcodes, 2D-barcodes, Bluetooth radio technology, or radio-frequency identification (RFID) tags.

Barcode systems

Barcodes or 2D-barcodes, such as QR and DataMatrix, enable semi-automatic data capture, through handheld or fixed readers. Printed barcodes can easily be attached to parts, products, packaging, or carriers and used to track and trace movements of parts. Since barcodes do not automatically transmit their location to any system, it is necessary to install or use readers at any point of data collection.

The term barcode is used for both 1D and 2D codes, although only the 1D versions have the characteristic "bars."There are different standards of barcodes. Figure 12.20 shows three typical examples: The Code 128 barcode that has been used in many industries for decades, the QR code, and the DataMatrix.

- **1-dimensional barcodes** are everywhere in retail. There are many different standards used for 1D barcodes. Barcodes were invented by Norman J. Woodland and Bernard Silver and patented in the US in 1951. But it was first in the 1980s, with the growth of global supply chains, that mass adoption of barcodes happened.
- **The Quick Response code, QR**, is a white square with black squares arranged on it. The QR code was first used by Japanese automobile supplier Denso in 1994 to allow fast scanning and tracking of items. Today, QR codes are very common in consumer advertising. They often contain data for a locator, identifier, or tracker that points to a website or application.
- **DataMatrix** symbols are usually squares, composed of square cells representing bits. Due to the ability to store up to 2,335 alphanumeric characters in a very tiny area plus excellent readability, a popular application for DataMatrix is marking small items such as small electronic components. DataMatrix is also used in industries that depend on excellent track and trace performance, such as the food industry, pharma, and aerospace. DataMatrix was invented by International Data Matrix, Inc. in 2005.

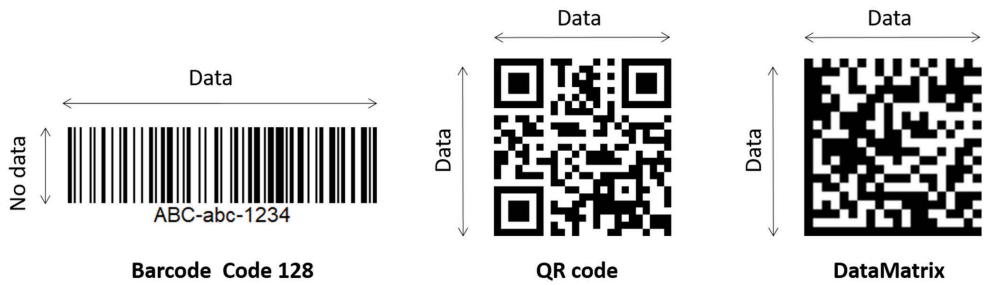


Figure 12.20 Examples of 1D and 2D barcodes

Compared to manual entry of data, barcode systems improve data accuracy and reduce the time needed for entry and error correction. Experienced typists make one error approximately every 300 keystrokes.

Barcodes have high read accuracy, are simple to manage, and cost-efficient. The downside is that all codes must be printed in high quality and scanned individually in a clean, well-lit place. Since they are printed, they are also prone to wear and damage. Also, barcodes are limited in what they can store.

The barcode is usually ink-printed onto an adhesive label, or laser-printed or jet-printed directly onto the part or product. Figure 12.21 shows an example of a DataMatrix printed onto a stamping tool produced by German company TRUMPF. The DataMatrix is printed at a place of the product that is not machine afterwards. This way, each product has a unique ID and can be tracked along the production process. The codes can be read and linked to transportation carriers' ID used in transportation. TRUMPF's customers can read the DataMatrix with an app and reorder the same type of stamping tool directly in the app.



Figure 12.21 DataMatrix printed onto a stamping tool and its base

A barcode system consists of the barcode, scanners, and backend IT software and hardware. A common 1D barcode printer for industrial use can typically cost around $1,000. Yet, small barcode printers can be bought for as low as $50 in retail shops. The same price range applies to scanners. In 2021, high-end scanners sold for about $3,000. Figure 12.22 shows two examples of readers and an industrial printer from the American company Zebra and a simple barcode printer from Taiwan-based Godex.



Figure 12.22 Examples of barcode scanners and printers (Photos: Zebra and Godex)

Radio-frequency identification

RFID offers the opportunity to read many units instantly. It does not require a visual line of sight to the RFID tag. RFID tags have also been called "smart barcodes." They are not really smart but they do not need manual operations to be read. Hence an RFID tag, when entering a reader zone, will automatically transmit "here I am, and here's what I have to say." RFID works by transmitting and receiving information via an antenna and a microchip (see Figure 12.23). The microchip on an RFID reader is written with whatever information the user wants.

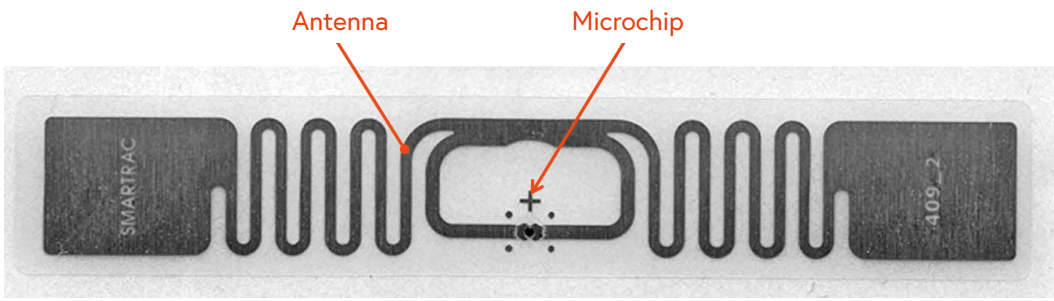


Figure 12.23 Anatomy of a passive RFID tag

RFID tags are either *active* or *passive*. The passive tags are most common, they use energy received from the reader to transmit a signal back to the reader. When they are not in the reader zone, they are not transmitting the data. These tags are very cheap – down to $0.10 per tag – and can be bought on adhesive rolls. An active RFID tag has a battery, which makes it constantly transmit the information. Because of the battery, it is larger than a passive RFID tag, more expensive, requires more maintenance, but able to transmit its signal over longer distances. Active RFID tags have a larger, more rugged design than passive RFID tags.

> **Stop and Think 12.4**
> You are likely to use RFID technology every day. Can you list five examples of passive RFID technology in everyday life? Do you use any active RFID tags?

In the early 2000s, RFID was foreseen to be heavily adopted in industry and radically change supply chain management. Despite plummeting costs and easy accessibility, the visions have not been fulfilled. Why?

Some important reasons for the unfulfilled promises of RFID are the following:

- **Infrastructure investments.** While tags are cheap, investments are needed for tag mounting, scanners, and IT. All hardware needs maintenance and data must be managed. Taken together, RFID costs much more than the 10–20 cents you pay for each passive RFID tag.
- **Reader accuracy.** While RFID tags have achieved high reader accuracy, it is not always perfect. Signal issues can occur when tags are broken, when signals from two or more readers overlap, or when there is interference caused by metal, water, or other magnetic fields in the surrounding area. If one in every thousand products is not tracked, it means data cannot be trusted.
- **Data security.** RFID tags do not distinguish between readers. Hence, the information can be read by almost anyone with a reader.

- **Implementation costs.** There are large experimentation, decision, and implementation costs related to RFID projects.
- **Maintenance and operating costs.** There are large maintenance and operating costs related to RFID implementations, including continuous training and software and hardware updates, replacement, and repair.

All these problems are eliminated if the 1D barcodes are chosen instead, which still make them the preferred choice in many cases in manufacturing.

To enable seamless track and trace in supply chains, systems need to be able to understand each other. The Global Trade Item Number (GTIN) enables this by communicating standard formats for barcodes, which can also be used for RFID information. Box 12.2 gives a short summary of GTIN.

Box 12.2 **Global Trade Item Number (GTIN)**

GTIN is an internationally recognized system for identifying products developed by the non-profit organization GS1. GTINs are a numerical representation of the product's barcode. GTINs can be 8, 12, 13, or 14 digits long. The types of GTIN that currently exist include:

- ISBN – International Standard Book Number (13 digit format)
- UPC – Universal Product Code (12 digit format) used primarily in North America
- EAN – European Article Number (8 or 13 digits)
- JAN – Japanese Article Number (13 digits)
- GTIN-14 – wholesale or multipack products (14 digits)

E-commerce companies such as Google and Amazon keep large databases containing almost all the GTINs in the world. To sell products through their platform services, strict adherence to GTIN is required. The use of GTIN helps them compare offers, improve visualizations, and understand demand and supply patterns. Currently, GTIN is used exclusively for bar codes, but it could also be used in other data carriers such as RFID.

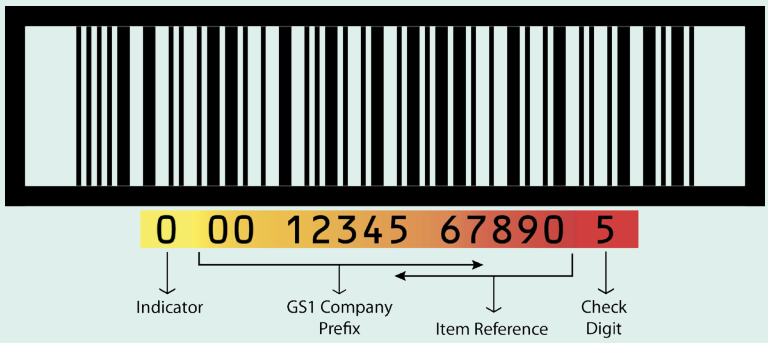


Figure 12.24 GTIN-14 bar code for wholesale (Credit: gtin.info)

12.7. **Third-party logistics**

Third-party logistics (3PL) involves the outsourcing of many logistics services beyond transportation, including fleet management, network design, Electronic Data Interfaces (EDI), materials handling, traffic management, and even supplier management. Started as an industry in the US, 3PL is now used by most manufacturing companies in the US, Europe, and Asia. Toyota uses it in North America, Europe, and Japan.

3PL is a special case of outsourcing, and manufacturers usually turn to it when the objectives of limiting their investments and reducing their costs take priority over control. Political considerations sometimes play a part as well, when a foreign company uses 3PL to engage and transfer know-how to local businesses, and thereby build a reputation for good corporate citizenship.

3PL providers do not make, consume, or sell materials but render logistics services to companies that outsource it; 3PL providers can take on most logistics tasks, but they should not be used where product knowledge is required, as is the case in kitting parts for assembly, checking the quality, or performing last-minute customization. The 3PL providers have their own resources and supplement them through partners as needed.

Here is an example of how a typical 3PL advertises its services:

- **Fleet management.** Providing and managing vehicles and drivers.
- **Engineering.** Designing logistics services for clients.
- **Material control.** Providing an information system for EDI, exception reporting, and receiving management.
- **Materials handling.** Sequencing, repacking, and consolidation through cross-docking.
- **Route and traffic management.** Dispatching drivers on routes, and providing a system for real-time tracking and tracing by customers.
- **Supplier management.** Working with suppliers on pickup times, conformance issues, and requirements for returnable containers and production materials, and training on lean manufacturing as needed.
- **Transportation management.** Transportation services supplement the 3PL's fleet with partner carriers like TST Expedited Services in Canada.
- **Warehousing.** Storing and protecting finished goods.

As soon as an acronym like 3PL gained currency, it was only a matter of time before "one-upmanship" led to the emergence of a "4PL," a term that has been both coined and trademarked by consulting firm Accenture. In Accenture terms, a 4PL provider assembles the resources, capabilities, and technology of its own organization and other organizations to design, build, and run comprehensive supply chain solutions. The difference with what we have been discussing above as 3PL is not obvious, and neither is the need for another middleman.

12.7.1. Late-stage customization

Leading logistics providers have started to offer activities that traditionally were done by the manufacturer. In particular, they offer late-stage customization. This postponement strategy makes a lot of sense when the market asks for a high level of cosmetic changes to any product. The typical example is packaging or labeling, but it can also be more profound changes. Box 12.3 shows an example from paint manufacturing, where the exact color of the paint obviously matters.

Box 12.3 Late-stage customization of paints

Jotun, a leading manufacturer of paints, does late-stage customization of B2B customer-specific colors in their warehouses. This way, Jotun only needs to produce base color (white) and color tinters, which are mixed in their own multicolor machines (Figure 12.25). It enables Jotun to mix exactly the color that the customer buys and still maintain a same-day delivery policy if ordered volumes are not excessive. Although Jotun is not using a 3PL vendor for this job, it could technically be done by one. When consumers buy customized paint colors in retail shops, they will typically be served by smaller multicolor machines.



Figure 12.25 Late-stage customization of paints in a factory in Abu Dhabi, UAE (Credit: Jotun)

With the advent of additive manufacturing – erroneously but commonly known as 3D printing –, 3PL vendors also see an opportunity to branch into spare-part production. From centrally located warehouses they can offer it as a service to customers and customer's customers. It is, however, the leap from logistics to manufacturing that is a great challenge. Box 12.4 discusses an example of a joint venture between Wilhelmsen and Thyssenkrupp.

Box 12.4 **Wilhelmsen and Thyssenkrupp offer marine spare parts to anywhere**

With the new opportunities offered by additive manufacturing, many traditional logistics companies are considering an entry into manufacturing. The idea is to take advantage of established global logistics networks and offer manufacturing of parts close to demand or at strategic logistics points where parts can be quickly shipped to any destination around the world.

One such initiative is a partnership between maritime company Wilhelmsen and the industrial engineering conglomerate Thyssenkrupp. The main aim of the partnership is to offer shipping companies and maritime equipment companies quicker replacement of spare parts including solving part obsolescence and part performance issues – independent of where in the world a vessel is located. The partnership being established in 2019, the Wilhelmsen Thyssenkrupp AM fulfillment platform serves more than 3,000 vessels from key AM locations worldwide. For example, the new venture has produced a cooling water pipe connector for Kawasaki Heavy Industries (Figure 12.26) – delivered with shorter lead time and improved part performance.



Figure 12.26 Cooling water pipe connector produced with additive manufacturing (Credit: Wilhelmsen)

The motivation for logistics companies to move into spare part manufacturing is obvious and the value-creating potential is tangible: Customers can *potentially* get products faster, cheaper, and in better quality. It is also more environmentally sustainable than sending parts over long distances. The ability to "print" and deliver spare parts on demand should theoretically result in radically more sustainable, effective, and efficient supply chains.

However, the challenges are not to be underestimated. A major hurdle is the legal right to produce parts (intellectual property issues), access to (the correct) digital product files, access to additive manufacturing capabilities, and coordination between sellers and buyers. Guaranteeing product quality can also be a challenge when relying on third-party additive manufacturing networks. Furthermore, additive manufacturing can only produce single parts and require pre- and post-processing. Whether spare part production will be revolutionized by additive manufacturing coordinated by logistics companies remains to be seen.

12.7.2. Strategic considerations for 3PL

For 3PL to be a viable option, it must be desirable for economic or political reasons, and companies that offer 3PL must be available in the geographic area. In principle, the outsourced services would not require product knowledge, but we have just discussed 3PL providers getting involved in additive manufacturing, which would obviously require product knowledge. The following sections elaborate on these points.

Economic and political issues

Wherever present in the world, for historical reasons, automobile manufacturing pays higher wages than other industries for similar work. In particular, materials handlers working in car assembly factories make substantially more than their counterparts at 3PL providers. In this case, outsourcing activtites to 3PL providers can be economically attractive. The same, however, is not true for every manufacturing industry. If you make food products, cosmetics or toys, the difference in wage structure may not be nearly as large and may not provide a similar incentive.

Even where there is no major wage difference, another source of cost reduction in the use of 3PL is the sharing of resources with other clients of the 3PL provider. This can be a dominant factor for small manufacturers whose activity could not possibly fill up a fleet of trucks and a network of warehouses or distribution centers.

Finally, even without wage differences and with a sufficiently high volume of activity, management still needs to assess the risks associated with owning and operating its own logistics resources through economic ups and downs. Even if the use of 3PL costs more under a given set of assumptions about the future, it may be prudent to pay a premium for another organization to take on the risk of overinvestment.

When Saturn, Ford, Chrysler, or Mission Foods use Penske Logistics, Schneider Logistics, or Ryder, there is no reason to think that they have any motivation other than economic. When Toyota adopts similar policies in North America and Europe, there may be the added motivation of nurturing a reputation as a good corporate citizen by engaging and training local businesses, thereby defusing any potential nationalistic backlash.

Availability of 3PL services

The availability of 3PL around the factory is not a foregone conclusion; 3PL, as well as other outsourcing efforts, are often limited by the surrounding infrastructure. Manufacturers who locate plants in isolated areas to take advantage of cheap real estate and tax incentives may place themselves in the position of having to perform all sorts of peripheral services themselves. The operator of a mine in the Andes must take care of providing panettone cakes to workers for Christmas as well as shipping out ore. On the other hand, a factory located in an area with a 100-year history of manufacturing can count on the availability of the services it may want to outsource.

Where road and communication networks are poor, and where there are borders to cross and tariffs to pay, 3PL services may be difficult to obtain but extremely valuable. For example, several US-based 3PL

providers have made an effort to reach remote places in Latin America. Even in industrialized areas, 3PL services may be only partially available. For example, there may be trucking companies offering no other services. A manufacturer with sufficient power can then encourage the formation of new ventures and train them to meet all 3PL requirements.

Product knowledge requirements

Third-party logistics providers being specialized in logistics cannot be expected to know their clients' products beyond handling requirements. They can transport parts in trucks, handle them on pallets, store and retrieve them in warehouses, and even transfer them between different types of containers without knowing anything more than outer dimensions, weight, need for refrigeration, orientation, or fragility.

On the other hand, some tasks that are usually considered part of logistics are better not entrusted to third parties because they require a deeper level of product knowledge, such as the following:

- **Kitting.** Kitting done by a 3PL provider is done too early, possibly days in advance of assembly, and subject to errors due to pickers not knowing the product. By contrast, kitting by water spiders is done a few minutes before assembly, next to the line, and by an operator who knows the product.
- **Quality checking.** The best time to check the quality of a supplier's parts is when picking them up in the supplier's shipping area. A driver working for the customer can be trained to recognize when a part has been mislabeled, but the most that can be expected of a 3PL driver is to check the paperwork and count bins or pallets.
- **Customizing.** Some suppliers provide last-minute customization services in or near customer plants. Since this is actually modifying the product it is strictly speaking production and not logistics, and it is also the last production operation before the customer sees the product, which does not make it a good candidate for outsourcing.

Logistics is a source of market intelligence

Since outbound logistics is the tail end of the order fulfillment process and deliveries trigger the collection of revenue, there should be no confusion within the company regarding the importance of this activity. If anything is done right except delivering the product to the customer, it's a disaster for the business. There is however one aspect of external logistics that makes it different from a mirror image of inbound logistics, and that is its role as a source of market intelligence. Some manufacturers realize that the physical organization of distribution provides direct first-hand access to market information.

It is not just downstream intelligence that gets lost with external logistics providers or 3PL, also upstream intelligence is given to the logistics provider. Compared to the use of one's own trucks and drivers to pick up parts on milk runs to suppliers, the use of 3PL implies a loss of valuable supplier information. Every visit by a truck to pick up parts from a supplier is an opportunity for the customer to communicate expectations and receive status information from the supplier, which is lost with 3PL.

> **Stop and Think 12.5**
> Today, few manufacturers organize their own external logistics. Why do you think this is the case?

12.8. **Supplier milk runs**

A supplier milk run is a scheduled pick up of parts from multiple suppliers in matching quantities, and a more sophisticated approach than a hub, reducing inventories of incoming materials, making lead times predictable even for items with variable consumption, smoothing the receiving workload, and providing an infrastructure for the transmission of pull signals.

Milk runs work primarily with local suppliers, but a local milk run can include a remote supplier who maintains a local warehouse. Clusters of remote suppliers can also be served by milk runs, with cross-dock or consolidation centers providing an interface to long-haul transportation by truck or intermodal railroad. Milk runs should not involve more than four or five suppliers, who are usually chosen for geographical proximity. If many are close to one another, then destination within the plant can also be used as a grouping criterion. Milk run routes can be planned using online mapping services but taking into consideration factors like border crossings, traffic and weather conditions, and road work in progress requires a more dynamic analysis that can be conducted with Microsoft Access or Excel or more advanced Operations Research software.

The supplier sequence can be turned into a milk run schedule, with arrival and departure times at each site. The execution of the milk run is tracked against the schedule, to refine it and to prevent schedule drift from affecting the supply of materials to the plant. The scheduled pick up of parts from multiple suppliers is designed to support a given production level. Instead of using the plant as a hub with trucks each shuttling back and forth to a single supplier bringing the same items by the full truckload, you have trucks making the rounds of multiple suppliers, returning empty containers and picking up less-than-truckload matching quantities of many different items, often adding up to a full truckload.

The differences between milk runs and direct transportation – where the plant works as a hub – can be seen from working out the implications of the patterns shown in Figure 12.27. Milk runs are a more sophisticated system than a hub and require more work to implement. However, their advantages often justify the extra effort.

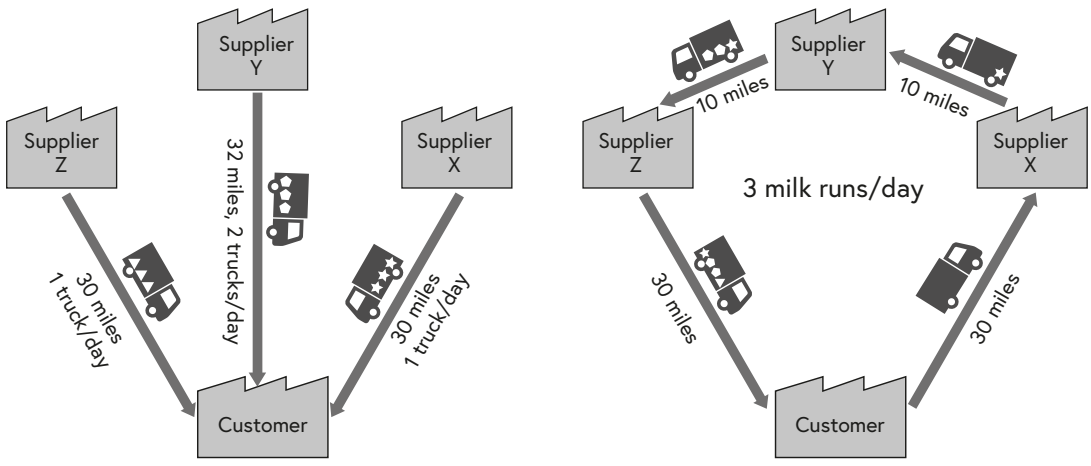


Figure 12.27 Supplier milk run versus the traditional approach

Because it only requires coordination between the customer and each supplier individually, the hub model is easier to set up than the milk run, which involves all the participants on each trip. In addition, the milk run may require more truck mileage to deliver the same volume of parts. In Figure 12.27, the milk run diagram assumes that the truck is large enough to hold 1/3 of a day's worth of X, Y, and Z. If that is true, the total truck mileage is about the same for both approaches.

The hub model diagram shows *four* truckloads delivered each day, requiring 2×30 miles+4×32 miles +2×30 miles = 248 truck-miles/day. Unless the X, Y, and Z parts have exactly the same dimensions and one truckload covers exactly, 1 day, a half day, and 1 day of production then the trucks will not all be full, and mix-loading trucks in the milk run mode may afford an opportunity to supply all three items in only three truckloads and the milk run requires 3×(30 miles +10 miles +10 miles +30 miles) = 240 truck-miles/day.

If the policy is to order only full trucks from each supplier, it is unlikely to result in exactly one truck/day or two trucks/day. The delivery frequency will then not be as shown in Figure 12.27, but instead vary for each item, resulting in alternations of congestion and low activity at receiving. If the trucks are truly full, then there is no way their content will fit into the *three* truckloads on the milk run side, and the customer will need either more milk runs or a bigger truck.

12.8.1. **The advantages of milk runs**

More information is needed to compare the performance of the two approaches depicted in Figure 12.27. Assume that the deliveries from the suppliers are done under the following conditions:

- Trucks going from the plant to the suppliers carry empty pallets and returnable containers.
- Suppliers X, Y, and Z each provide one item going into the same assembly, at the rate of one per unit for X and Z, and two per unit for Y.
- Assembly proceeds at a constant pace.
- The same truck is used for all three suppliers, so that, in the hub mode, the trips to the three suppliers are staggered throughout the day.

Using this example, milk runs offer the following four advantages:

1. Reduced inventory
2. Predictable replenishment lead times
3. Inventory visibility
4. Improved supplier communication

Reduced inventory

Figure 12.28 compares the impact of the two approaches on the inventory of the corresponding items inside the customer plant. The saw-tooth pattern, with a smooth linear decrease between deliveries, is generated by constant production at assembly. The safety stock serves to protect production against fluctuations in travel times.

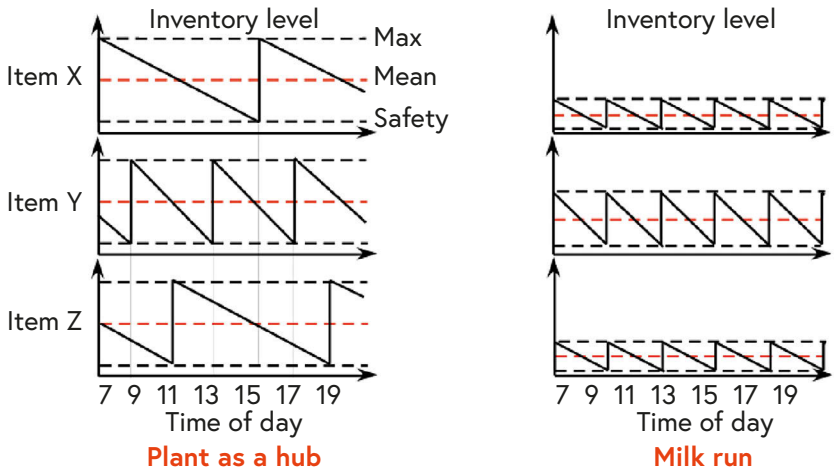


Figure 12.28 Inventory levels with hub versus milk run

For items X and Z, by multiplying the delivery frequency by 3, the milk run approach reduces the amount of inventory to 1/3 of the hub level for items X and Z, and 2/3 of the hub level for item Y. Likewise, milk runs involving 5 or 6 suppliers stand to reduced inventories by 80% or more. The number of pallet locations, pallets, and returnable containers required are proportionately reduced. The frequency of use of returnable containers is closely tied to their economic justification. Since they only become cheaper than disposable containers after about 20 uses, the faster they accumulate these 20 uses, the earlier they pay for themselves.

Predictable lead times

The value of milk runs is not limited to items of constant consumption. A milk run is, in fact, a predictable transportation channel. A regularly scheduled truck arrives every two hours or every three hours; its content may vary but its arrival time doesn't. In practice, milk run trucks arrive at each site on their route within a 15-minute window.

Figure 12.28 shows *three* items with constant consumption. Real manufacturing situations, however, involve not three but fifty, five hundred, or even 5,000 items with different delivery frequencies. Getting them delivered individually by the truckload then becomes a complex problem, regardless of whether the customer owns the truck or third parties are involved. Then milk runs *reduce* complexity and make it possible to order quantities that are less than truckloads without generating enormous transportation costs.

The advantages of milk runs are even more compelling when item consumption is variable, because, in a hub system, trucks cannot be counted on to be available whenever needed and can only be used economically for sufficiently large quantities. This adds a layer of unpredictability in lead times that the procurement process doesn't need.

The stocks of materials and components inside the plant are sized so as to be just large enough to support consumption until the next delivery. Consumption rates are based on production plans that are internally developed to meet the demand. The replenishment lead times, on the other hand, depend on the supply network, and their variability is one of the most common complaints of production control managers.

Inventory visibility

With the staggered delivery pattern of a hub system, as shown on the left side of Figure 12.28, the amount of materials of any item on warehouse shelves provides little information. Racks can be full of X's and nearly empty of Z's without it signifying any abnormality. By contrast, milk runs deliver matched quantities of multiple items at the same time, so that, in normal operations, the racks of the different items are full, half-empty or nearly empty all at the same time. As a result, any visible imbalance between available quantities among these items is an immediate cause for concern.

Improved supplier communication

Using milk runs means visiting suppliers more often, which in turn provides numerous opportunities for communication with the supplier about delivery or quality. How well both sides are able to take advantage of these opportunities, however, depends on how the milk runs are set up. If the pickup operations are handled by a third party, then the drivers will have no particular knowledge or concern about the customer's products. On the other hand, if the drivers are employees of the customer plant, their role can be enhanced to include verifying at least that they are given the right parts in the right quantities and passing on replenishment orders together with empty containers.

12.8.2. **Where milk runs don't work**

As stated earlier, one of the key characteristics of logistics is that it is not a one-size-fits-all approach. Though milk runs are powerful, they are not applicable to the procurement of every part from every supplier. Milk runs are primarily used for items that are frequently used in moderate quantities and come from suppliers that are within tens of miles of the plant and of each other. We shall discuss some adaptations of the concept to clusters of remote suppliers – that is, groups of suppliers that are close to one another but remote from the customer plant.

The following are some circumstances where the milk run concept is not applicable:

- **Very high-volume runners.** An item is needed in multiple truckloads every day. Then it makes sense to dedicate trucks to it, and possibly to have the vendor take charge of maintaining the supply in the plant.
- **Strangers.** An item is required only sporadically, in small quantities, from a supplier who does not provide any regularly used items. Then common carriers can be a better solution.
- **Distant suppliers.** The supplier is geographically far from any other that could be combined with it in a milk run route. Then the cost of making a truck that is empty or loaded with other items takes a side trip through that supplier's location cannot be justified.

The determination of whether a group of suppliers is a milk run opportunity requires a quantitative analysis that is specific to the industry and the companies involved. Expressions like "tens of miles" or "every day" apply in the automotive industry. In aerospace, by contrast, milk runs can be justified for the procurement of items costing in the tens of thousands of dollars per unit, from a cluster of suppliers that is hundreds of miles away and needed twice a week.

Despite its benefits, supplier milk runs are not commonplace. It is much more common to outsource logistics to third-party logistics providers and not engage in the details of pick up. Victor Fung,[3] from Hong Kong's apparel distributor Li & Fung, reported collecting assortments ordered by customers from up to 10 manufacturers by milk runs. In this application, the materials are needed only for shipping, and therefore the timing requirements are less stringent than for assembly.

What kind of truck for a milk run?

The use of trailers or semi-trailers provides some flexibility. If need be, a cab can be unhitched from a trailer at a receiving dock or even in the parking lot, and immediately hitched onto another trailer pre-loaded with empty pallets, returnable containers, and kanbans for another milk run.

Most trucks are loaded and unloaded from the back, through docks that, in the US, are 9 to 12 feet wide, raised 42 to 60 inches above the ground. These trucks can only be unloaded in the last-in-first-out (LIFO) mode. If they contain materials from 5 suppliers, then the contribution from the first supplier on the run is at the front of the truck and only available after the materials from the following four are unloaded. The unloading speed is also limited by the width of the truck. The much less common side-loading trucks provide way better access to their loads but are often incompatible with facility ramps. In particular, materials handlers can unload at the same time materials from different suppliers bound for the same destination inside the plant.

The differences between the two approaches are shown in Figure 12.29.

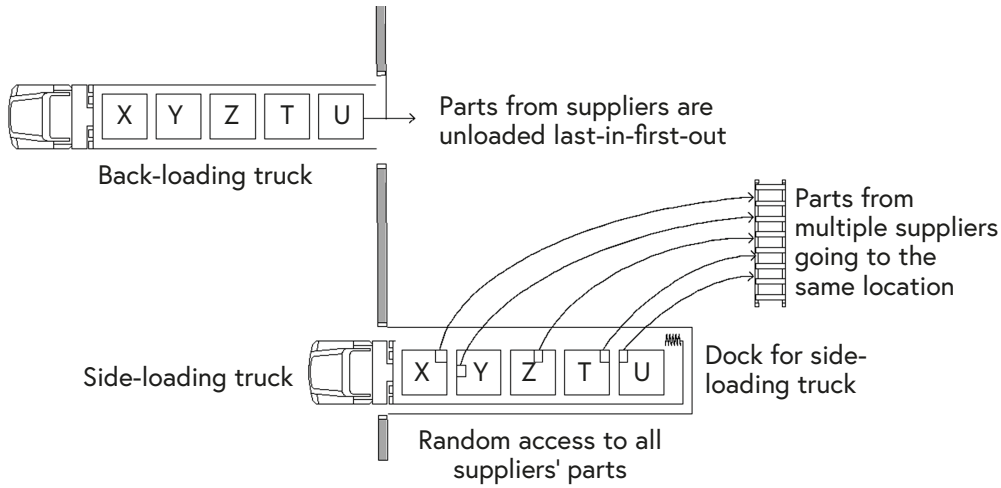


Figure 12.29 Back-loading versus side-loading trucks

The unloading time can also be reduced by applying more handlers concurrently. Side-loading trucks, however, remain a minority for two reasons:

1 They require docks that can be up to 50 feet long and practically have to be inside the building. This is not a common configuration and would have to be present at all plants along the milk run.

3 Margretta, J. (1998). Fast, global, and entrepreneurial: Supply chain management, Hong Kong style. *Harvard Business Review*, *76*(5), 102–115.

2 The load/unload side of the truck is usually closed by some form of a curtain, as shown in Figure 12.30, which does not provide as mechanical, thermal, or humidity protection to the freight as the solid walls of back-loading trucks.



Figure 12.30 Side-loading truck with curtain closure

Stop and Think 12.6
What type of products do you think are unsuitable for transportation with side-loading trucks?

12.9. **Consolidation centers**

A consolidation center is a facility located near a manufacturing plant that receives components and parts from many suppliers and delivers them to the plant as shown in Figure 12.31. It is usually operated by a separate company, in which the manufacturer may or may not own equity.

In general, adding an intermediary between suppliers and users does not enhance a supply chain, and it is, therefore, surprising that leading car makers still do it. They have many different motivations. A consolidation center can shield the factory from dealing with overseas suppliers with lead times in months. The consolidation center is then a domestic supplier, working with the plant as if it made its own parts day-by-day.

The complexities and risks of working with remote suppliers and four-month lead times do not disappear, but, rather than dealing with them, the customer plant prefers to pay the consolidation center to

do so. The plant staff has its hands full keeping assembly lines and machines running and can do without the additional workload.

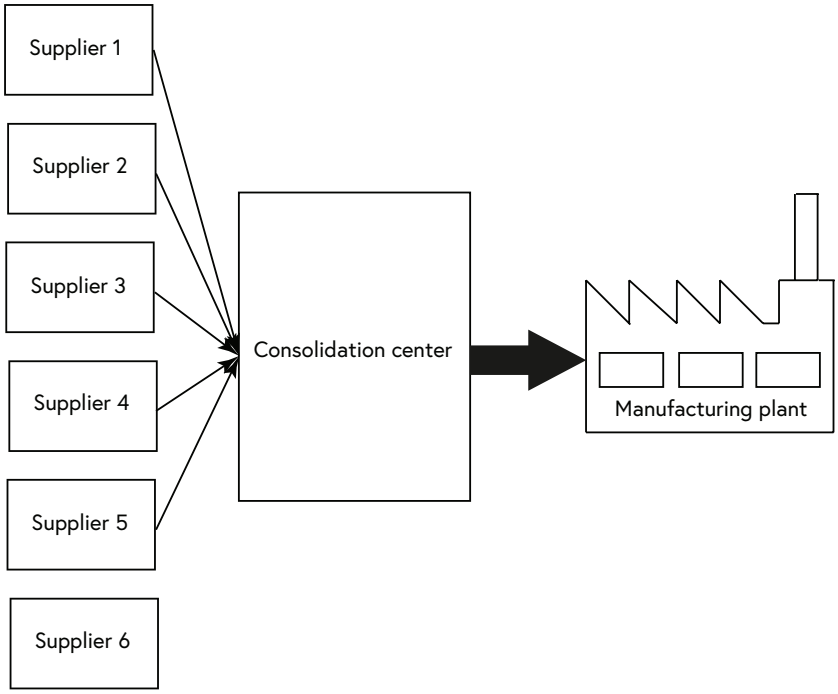


Figure 12.31 Illustration of a consolidation center

A second, similar motivation is to insulate the plant from domestic suppliers who won't work with kanbans or returnable containers. These suppliers should convert or be weeded out but it takes time, and, meanwhile, their products still need to be delivered to the plant. Again, the consolidation center relieves the plant from having to accept large, sporadic deliveries of inappropriately packaged parts.

While the first two reasons for using a consolidation center apply to other industries as well, labor cost reduction is a strong third one that is special to the automobile industry. In every country where it exists, the automobile industry is and has been paying high wages to its operators, whether they work on the assembly line or in the warehouse. Outside of it, a consolidation center can recruit warehouse personnel for half or even one-third of a car assembler's wages. It cannot do all the materials handling for the manufacturing plant, but what it can do means it can make a profit while saving money for the plant.

12.9.1. How a consolidation center works

The flows of materials into a plant that uses a consolidation center are as shown in Figure 12.32, and discussed in the following paragraphs.

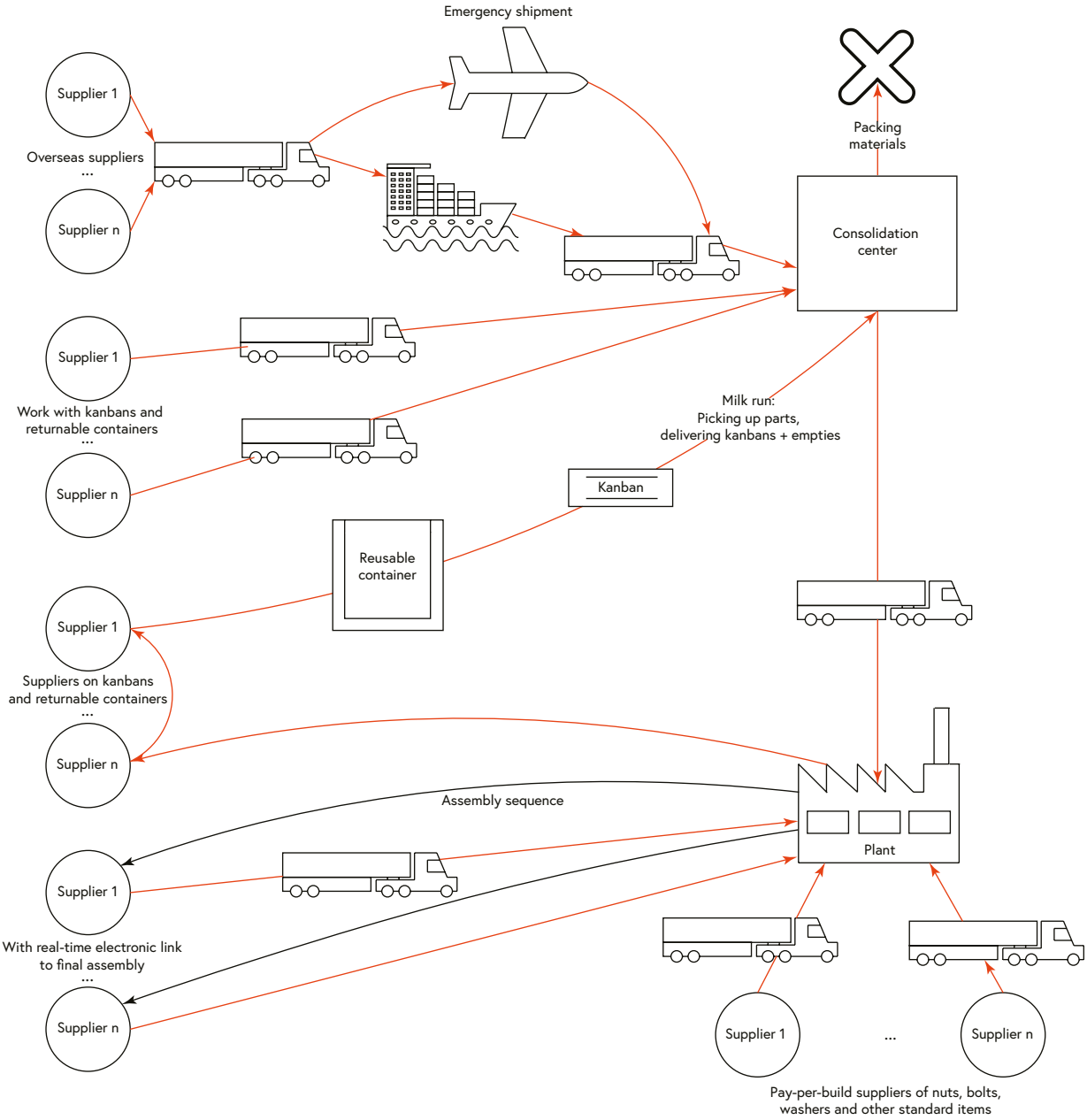


Figure 12.32 Material flows to an assembly plant: Direct supply (bottom), milk run (middle), and consolidation (top)

Not every item needs to go through the consolidation center. Let us assume that its purpose is to insulate the plant from overseas suppliers and from domestic suppliers who do not deliver at the frequencies, in the quantities, and in the containers that the plant wants. Then the suppliers of all other items should deliver them directly to the plant.

In particular, the following type of items should not go through the consolidation center:

- **Body-on-sequence items.** The supplier is notified electronically when each unit starts final assembly and starts making it then in order to deliver it to the appropriate assembly station just when the unit reaches it. This is used in the car business for such model- and option-dependent items as seats and body upholstery.

- **Pay-per-build items.** The supplier is responsible for maintaining stocks at or near the line side and is paid based on the bill of materials for finished products leaving the assembly line. In-car assembly, it is used for nuts, bolts, washers, and similar standard items. The suppliers of these items may in fact consolidate inputs from second-tier suppliers, but that is not the same thing as going through the consolidation center of the plant.

Over time, the plant should attempt to source all items locally, and the populations of both overseas suppliers and uncooperative domestic suppliers should wither. The consolidation center should, therefore, be needed most when the plant is new, and less and less as the plant matures. No matter how far we look, however, not every supplier will have a factory within 10 miles of the plant, and those who don't will be under pressure to open at least a warehouse there. The consolidation center can find a long-term mission in serving as a common warehouse for a group of these suppliers. An example is given in Box 12.5.

Box 12.5 **Mercedes Benz consolidation center in Speyer**

Consolidation centers do not always serve one factory. With the rise of global logistics the concept has expanded to serve multiple plants – often whole global production networks. One example is Mercedes Benz's consolidation center opened in Speyer, Germany, in 2016. The idea is to give central responsibility for efficient and flexible worldwide logistics. The center in Speyer consolidates parts and materials from European suppliers and supplies the Mercedes-Benz assembly plants in China, the U.S., and South Africa. The center's job is to secure the on-time supply of materials to these sites, while reducing overall inventories in the network. Parts are consolidated and shipped in containers via sea freight. The center, which has 80,000 square meters of hall space and 21,000 square meter container yard, is operated by third-party logistics provider Syncreaon under the responsibility of a Mercedes-Benz special supply chain team. The consolidation center employs around 500 people as of 2021.



Figure 12.33 Mercedes Benz consolidation center in Speyer, Germany (Credit: Daimler)

12.9.2. **Physical organization of the work**

The consolidation center receives large shipments, breaks them down into the smaller quantities that the plant needs, disposes of the supplier's transportation crates, boxes, and bags, place the parts in the plant's reusable containers and delivers them either to a plant's Receiving department or directly to the line side. Because it breaks large quantities into small ones, the consolidation center holds substantial inventory and needs a warehouse.

A concept for a consolidation center is shown in Figure 12.34. It is radical in the sense that it uses trailers for storage and is free of forklifts, and there are many more alternative ways of doing this, using, for example, side-loading trailers and flow racks that hold entire pallets.

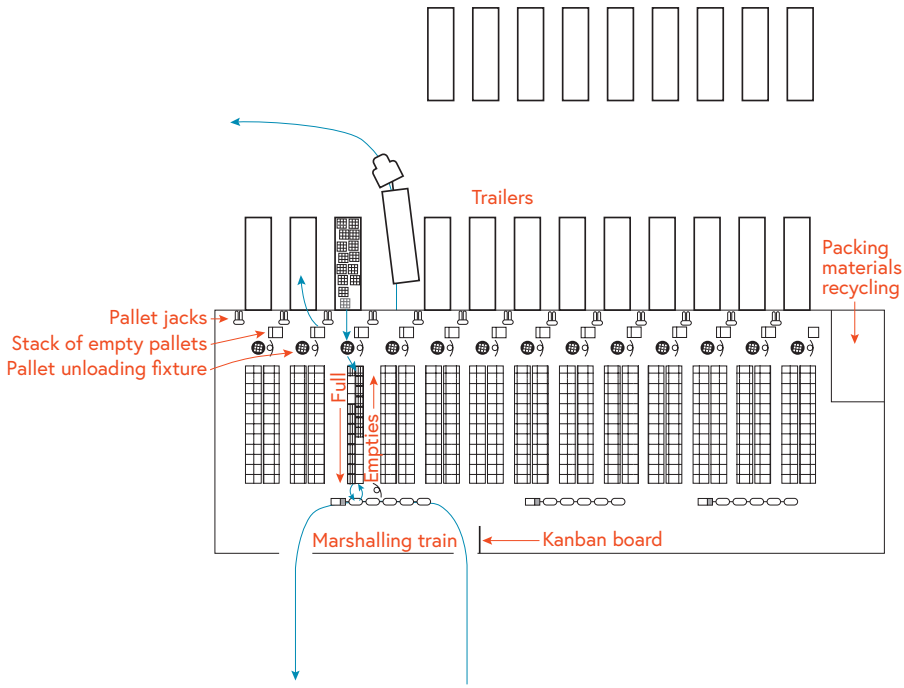


Figure 12.34 A consolidation center concept

Before considering the details of the concept in Figure 12.34, we should recognize its limitations:

- It is intended for parts like automotive pumps, compressors, or headlights that fit in standard containers that can be transported on pallets. It wouldn't apply to 2-dimensional parts like sheet metal or 1-dimensional parts like bars or wire harnesses.
- It is intended for parts that have been packed so that they will not be damaged by exposure to the humidity and temperature changes they may experience in a trailer over a few weeks. Parts shipped by sea certainly meet this condition. Parts from remote domestic suppliers may or may not.
- It requires the same item to always be delivered to the same loading dock. If the dock is occupied, there should be no opportunity to unload at another.
- Items are grouped by destination on the assembly line, to support contiguous if not one-stop shopping for tugger trains. With reliable suppliers, this is appropriate, but, for questionable suppliers, you may prefer to organize stores by source, for easier monitoring.

To avoid multiple handling of pallets, all stocks are kept in trailers. As long as the orders placed four months ago match today's demand, for each item, there is one trailer being unloaded and some trailers on the road in various phases of fetching the next load. If too many orders were placed for today's consumption, then full trailers accumulate in the parking lot. The economics of this option depends heavily on the relative costs of real estate and construction in the area since a classical warehouse will have a smaller footprint for the same amount of storage space. The trailers used can be rented. As a storage system, they have the advantage that their number can easily be adjusted up or down.

Each loading bay has a pallet jack the operators use to move pallets from trailers one at a time, onto a lift-and-turn table to present the boxes in a consistent, ergonomically adequate position. If the flow rack for full item-specific returnable containers is not full, the operator then places the parts into the next container picked from the empties' flow rack.

For pallet unloading, tables that rise and rotate or tilt are available from several vendors as shown in Figure 12.35. Tables that both rotate and tilt, on the other hand, do not appear readily available. If only one of the functions is available, then rotation is more important than tilting. These devices may also need to be recessed into the ground for loading by pallet jacks.



Figure 12.35 Lift-and-tilt versus lift-and-rotate tables

The operator stacks empty pallets for reuse, along with any skids or item-specific, blow-molded separators used in transportation. Disposable packing materials go to recycling on an underground or overhead conveyor *not* shown in Figure 12.34.

On the opposite side of the flow racks, Figure 12.34 shows tugger trains delivering empty containers and picking up full ones. If the consolidation center is in the same industrial park as the plant, the tugger trains can go directly to the lines to deliver the parts. Otherwise, the full tugger trains can roll onto trucks and roll straight off the trucks onto the shop floor at the plant.

The kanban board is useful if some of the kanbans picked up at the plant are not for immediate delivery. For example, they may specify that delivery should take place in the milk run after next, which may be done by a different train. The driver, therefore, places the cards on the board in the corresponding column and picks up the cards that are due on the current run.

To increase the number of items available for picking at a single train stop, one can do the following:

- Use the narrowest possible flow racks.
- Place the return rack for empties above the rack for full containers rather than side-by-side.
- Stack more than one level of flow racks.

- Use mirror-image flow racks with docks on both sides of the building and have the train go down a central aisle.

From a strict logistics point of view, the best location for the consolidation center would be inside the plant building. Transportation from the consolidation center to the line side could then rely on the same tugger trains or carts used for other items, and the weather would not matter. If, however, the purpose of the consolidation center is to insulate the plant from certain supply problems, collocating the two in the same building may not be the best way to achieve it. At the management level, proximity and the sharing of food service facilities may make it difficult for plant materials managers to refrain from intervening in the management of the consolidation center. In addition, close daily contact between operators who make different wages for similar jobs is bound to create tension.

12.9.3. Three functions a consolidation center should not perform

There are limits to what can be reasonably asked of a consolidation center. The following are three examples of work that, for different reasons, are better kept in-house.

1. Kitting. The consolidation center should not be asked to pick kits for production, for several reasons:
   - To pick kits, you need pick lists based on the product's bill of materials, and sharing a bill of materials with the consolidation center means not only letting out key proprietary information but keeping it up to date. Whenever engineering changes are implemented in the plant, they must also be at the consolidation center. The corresponding exchanges of data are much more complex than if the consolidation center only deals with individual items.
   - The best time and place to pick a kit is next to the assembly line and just before it is assembled. The pickers can be experienced assemblers who will do a more careful and thorough job as a result of knowing what each part is for. Missing or defective components can be easily replaced, and there is no need to manage and transport a stock of kits.
2. Incoming quality assurance. Again, the issue here is the information that consolidation center employees need to have. They would have to be trained both on the plant's QA methods but also on the characteristics of the parts. This is better done inside the plant – without sharing technical information with outsiders – and where engineers are present who can analyze problems as they are detected.
3. Sorting empty boxes and dunnage. The temptation is great on the production shop floor to throw all the empty boxes and dunnage together at random and count on the consolidation center to sort it out. After all, consolidation center labor is cheaper than assembly plant labor, and therefore why not do it? It is not truly a transfer of work to the consolidation center but the creation of work that otherwise wouldn't need to be done. The empties on the shop floor are found where the parts have been consumed. They come out automatically organized by item, operation, and product line. By scrambling the boxes, production is actually destroying a structure it already had and that the consolidation center must later reconstruct. In this case, it is more economical to organize the pickup of empties by item.

Summary

The most essentials points covered in this chapter were:

1. External logistics is about managing inflows and outflow of the factory.
2. Logistics has become a truly global activity, partly due to containerization.
3. Packaging is essential for external logistics and must be carefully designed and adapted to the product shipped.
4. Part and container tracking allows managing flows and knowing the status of shipments and the whereabouts of parts.
5. Outsourcing external logistics to 3PL vendors has become commonplace but is a strategic decision that shouldn't be taken lightly.
6. Logistics companies are entering manufacturing with late-stage customization and additive manufacturing solutions.
7. Milk runs are often the most efficient way to manage flows from local suppliers to an assembly factory but is more complex to implement than direct shipments.
8. Consolidation centers help manufacturers reduce the complexity of incoming material flows.

Test yourself

1. Under what conditions would you prefer returnable containers to disposable containers?
2. What is dunnage and why does it matter?
3. What is the purpose of Incoterms?
4. What is third-party logistics (3PL)?
5. What is late-stage customization and why do 3PL vendors offer it?
6. What are the differences between barcode systems and RFID?
7. What are the advantages of supplier milk runs?
8. What is the role of a consolidation center and what are its advantages?
9. Why do manufacturers often not operate their own consolidation centers?
10. Make a rough sketch of the material flows in a consolidation center.

Further reading

Baker, P., Croucher, P., & Baker, P. (2017). *The handbook of logistics and distribution management: Understanding the supply chain*. Kogan Page Publishers.

Choi, T. Y., Li, J., Rogers, D., Schoenherr, T., & Wagner, S. (Eds.). (2022). *The Oxford Handbook of Supply Chain Management*. Oxford University Press.

David, P. A. (2013). *International logistics: The management of international trade operations*. Cicero Books LLC.

Ferdows, K. (1997). Made in the world: The global spread of production. *Production and Operations Management 6*(2), 102–109.

Kim, K. W. & Günther, H.-O. (Eds.). (2007). *Container terminals and cargo systems: Design, operations management and logistics control issues*. Springer Verlag.

Lowe, D. (2005). *Intermodal freight transport*. Elsevier Butterworth-Heinemann.

Novack, R., Gibson, B., Suzuki, Y., & Coyle, J. (2018). *Transportation: A global supply chain perspective* (9th ed.). South-Western College Publishing.

13 Warehouse Management

Most goods do not go directly from suppliers to production and consumption. Instead they wait somewhere, with the supplier, the manufacturer, the customer, or intermediaries like a distributor, a retailer, or a transportation company. Warehouses are buildings or sections of buildings used to store and retrieve goods.

Some products or materials undergo curing, fermentation, or aging processes that require them to be stored under specific conditions of temperature or humidity for periods of time varying from hours to years – for example, cheese, stockfish, and wine – but this is not the kind of storage we are concerned with in this chapter. This chapter looks strictly at storage that is necessary to buffer capacity from demand. Particularly in make-to-stock manufacturing, warehouse efficiency can be a source of competitive advantage.

A warehouse must be managed. This chapter explains the basics of warehouse design and operation. It also includes capacity decisions, layout and design of racking and shelves, picking technologies, organization, shipping processes, and automation.

The focus in this chapter is on what a manufacturer must know about warehousing. It is not about managing the mega warehouses and fulfilment centers of wholesalers and retail companies like Amazon, IKEA, or Walmart, nor is it about managing logistic hubs such as the cross-dock and consolidation centers of FedEx or DHL. However, these companies are logistics specialists and serve as best practice examples for manufacturers.

Learning objectives

After working through this chapter, you will be able to do the following:

1. Explain what inventory is and why it is needed.
2. Know about Warehouse Management Systems.
3. Plan the layout and logistics in a warehouse.
4. Choose appropriate technology for a warehouse.
5. Implement visual control in a warehouse.
6. Know the basics of put-away and picking strategies.

DOI:10.4324/9781351110310-17

13.1. The need for inventory

Some stocks of raw materials, work-in-process, and finished goods are as necessary to production as blood to the human body. When walking through a shop floor, a rack full of materials is not by itself an unmistakable sign of waste. If, however, there is no sign indicating its purpose, and operators, managers, and engineers cannot explain why it is there, then chances are it is not needed and you have found inventory waste. Other telltale signs include paperwork indicating that the material has been in that location too long, as when there is thick dust over the containers.

The need to retrieve, store, and manage raw materials, WIP, and finished products arises where there is a mismatch between incoming and outgoing quantities:

- **Long throughput times, short response times.** Lead times for purchased materials plus manufacturing throughput times are longer than the time the customer is willing to wait for product delivery. This requires a make-to-stock production strategy, which holds inventory by definition.
- **Trucks in, pallets out.** An item is delivered from a supplier by the full truckload and consumed in production one pallet at a time at a regular rate. This pattern generates up to one full truckload of raw material inventory for this item, with an average of half a truckload.
- **Totes in, pallets out.** A production line makes one unit of product at a time, accumulating the finished units into a tote at the end, and the totes are palletized for shipment. As above, there is an average of half a tote at the end of the line and half a pallet in stores, making work-in-progress inventory inevitable.
- **Pallets in, trucks out.** Up to one truckload of pallets of finished goods accumulates in the shipping staging area.
- **Single-item trucks in, mixed-loads out.** In cross-docking operations, full truckloads from different sources are rearranged and shipped out by destination.
- **Fixed rate production to fluctuating demand.** A production line may make consumer goods whose demand fluctuates. Or it may make components for industrial customers whose own lines work at a steady pace but who use ordering software that introduces artificial fluctuations. In either case, producing at a steady pace and using finished goods stocks to absorb fluctuations is sometimes preferable to maintaining the capacity to handle peak demand at all times.
- **Fixed rate production from fluctuating supplies.** Most production organizations are unable to deal exclusively with suppliers who always deliver perfect quality parts, in the right quantities, and right on time. A special case is the manufacture of seasonal harvest, like berries, fruit, and grain. And even perfect suppliers can be delayed by snow storms, floods, or traffic accidents. Safety stock is impossible to hold for custom items and too expensive to protect against events like the supplier's plant burning to the ground or even going on strike. On the other hand, it would be unreasonable not to protect production against the possibility of a 30-minute delay in truck arrivals.
- **Shipping to a steady demand from an unstable production line.** While this is a situation that lean manufacturing is intended to avoid, it is not achieved instantly, even when diligently pursued. In the meantime, stocks of finished goods may be needed to protect shipments against failures of the plant's own ability to produce, be they due to component shortages or internal process problems.
- **Bulk-buying discounts.** Taking advantage of volume discounts or suppliers' sales campaigns.
- **Holding safety stock.** Even when there is no apparent mismatch between supply and demand, manufacturers choose to keep safety stocks of material and products "just in case" something happens in the supply chain. Safety stocks generally increase with increasing distance to suppliers and markets.

13.1.1. Is inventory a waste?

As shown in Figure 13.1, one of the best known metaphors in manufacturing is that of inventory as water covering up problems in the form of rocks on the bottom. The word "inventory" figures prominently in Taiichi Ohno's list of 7 types of waste. As a consequence, when discussions of inventory or warehouse management arise, several participants immediately say that, since "inventory is waste," it should just be disposed of. But Ohno never said that all inventory was waste, only that *excess inventory* was a type of waste that should be *targeted for reduction*. His point was to draw attention to the improvement opportunities lying in factories' tendency to hoard more than they need. Lower the inventory, or the "water level," and you expose the "rocks" or the problems to be solved.

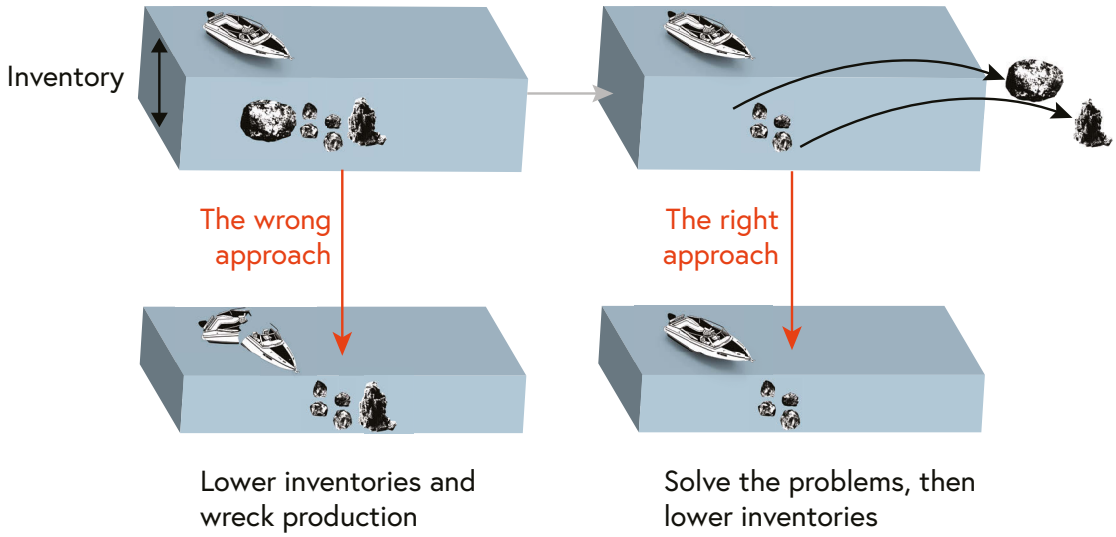


Figure 13.1 Lowering inventory to expose problems

It is true that mass producers have used high levels of inventory as a means of working around problems they had no plans to solve. However, if all you do is lower the water level, as shown on the left side of Figure 13.1, the rocks come closer to the surface and wreck your ships. Since this is easy to do, many companies have tried this and missed shipments as a consequence. Managers who have tried this are likely to discard the idea of low inventories and instead add more, "just in case."

The right approach is to *dovetail the reduction in inventory with the solution of the problems*. No matter how diligently pursued, the elimination of shortages and the improvement of internal processes are not instantaneous, and therefore the reduction or elimination of finished goods stocks should not be either.

The lot-size adjustment buffers described above are inevitable, but they are small, and of sizes that can be calculated easily. Sizing safety stocks is a more difficult challenge. The lean approach is based not on attempting to maintain large stocks of all needed items but on vigilance, detecting anomalies early, and responding quickly. Again, prudence is required in determining what is excess inventory.

If a plant consumes small quantities of an alloy with a single supplier worldwide, then management has no choice but to accept the supplier's terms, which may include a six-months order lead time and a minimum order quantity that covers two years of consumption. On the other hand, if an item is a

commodity that can be delivered overnight from a local supplier, then it is difficult to justify holding a week's worth of it.

13.1.2. Finished goods stores

If all products are made based on pull signals from customers, then all units can be shipped immediately and there is no need for stores of finished goods. It is, however, a big "if." When factories get to this point, it is frequently because, in fact, their customer is the Sales department. The Sales department issues orders to the factory that commingle actual customer orders with orders based on its own forecast and promotion plan, and owns the finished goods when they leave the factory. In this case, the factory and the manufacturing organization as a whole don't have to worry about finished goods inventories, but the Sales department and the company as a whole do.

Body-on-sequence suppliers to the automobile industry genuinely have no finished goods inventory. They do not start making a product unit until they receive a signal from the car assembly plant that the car it goes into has started on the first station of final assembly and ship it on the next milk run. Box 13.1 gives the example of Injex Industries that produced door panels to Toyota's NUMMI plant.

Box 13.1 **Injex, a body-on-sequence supplier to NUMMI**

Injex Industries, of Hayward, CA, was a body-on-sequence supplier of door panels to NUMMI in Fremont, CA, since 1993. Finished goods inventories of door panels at Injex went from a 14 days' supply in 1985 to 2 days in 1990, 4 hours in 1992, and *none* as of 1993. While this obviously required a sustained effort, it also shows a transition period of 8 years, during which the company had finished goods stores to look after.

By the 2000s, Toyota considered Injex a "showcase supplier," to which it sent other suppliers to learn. NUMMI was Injex's dominant customer but was dissolved as a Toyota-GM joint venture in 2009 when GM went bankrupt. Injex ceased operations in 2010. Other Toyota suppliers using body-on-sequence include seat maker Johnson Controls.

13.1.3. Consumables, spare parts, tools, containers, documents, etc.

In addition to materials that go into products the plant also needs to store and retrieve other items:

- **Consumables.** Although rarely given much attention, a factory can carry a significant inventory of consumables, such as gloves, paper towels, cardboard, or label rolls.
- **Spare parts for facilities and equipment maintenance.** Like production materials, most spare parts make a one-way trip to where they are used, but their flow is irregular. A steel mill may keep a spare pump in stock for years, as insurance against emergencies. The management of such stores involves periodically testing such devices.
- **Tools, fixtures, jigs, dies, etc.** Such devices are mounted on machines during setups, then cleaned and returned to stores after use, periodically refurbished, and eventually replaced. They are checked out and returned like library books rather than consumed like parts. Being managed directly by Production, these stores are outside the scope of this discussion.

- **Returnable containers.** Like tools and dies, they are recirculating resources, but, unlike them, they are managed by Logistics.
- **Test samples.** Test samples sometime have to be stored for many years in industries like pharmaceuticals and chemicals. Although samples are small, they can amount to significant inventories over time.
- **Documents.** Manufacturers frequently need to store paper travelers and documentation of quality checks for future trace. As far as possible and allowed, document storage can be digitized, reducing a library of physical information to a hard drive, with backups.

13.2. Warehouse planning and design

The stereotypical image of a warehouse is a large space filled with inventory racks. That is, however, not all there is to a warehouse. There are many different ways to lay out and organize a warehouse. A warehouse starts out as an empty space, in three dimensions. You allocate sections of its floor to different classes of items, depending on their physical characteristics and usage patterns, and you decide how much of the vertical space above the floor you use.

The size, weight, environmental requirements, and expiration dates of materials constrain your choices. Some items are used daily, others sporadically. If you limit the height of storage devices to what operators can reach while standing on the floor, you will have high visibility but will leave many cubic feet overhead empty. If you fill up the vertical space with racks, you need to invest more, particularly in high-reach materials handling equipment, and you lose visibility.

Figure 13.2 contrasts the design choices made by Amazon and IKEA. Amazon sells many small items picked manually and shipped to consumers; IKEA sells furniture in flat packs that consumers pick directly onto shopping carts from the bottom shelves, while the upper shelves contain additional inventory that is lowered by forklift to replenish the bottom shelves. IKEA also have dedicated distribution centers that serve the retail stores. A typical IKEA retail store warehouse would be organized with 5–6 height levels of single or double deep shelves.



a) Amazon stores goods on low shelves (Credit: Álvaro Ibáñez, CC BY 2.0)



b) IKEA uses the whole vertical for shelf space (Credit: SmitBruins/Twenty20)

Figure 13.2 Warehouse design: Amazon versus IKEA

The key question is whether the design fits your actual mode of operation. The prudent approach is to start cheap and simple, then invest in refinements and automation as operations reveal the needs. There is no one-size-fits all to warehousing.

13.2.1. Tailor the warehouse to the need

As in many other aspects of manufacturing and logistics, underperformance in storage and retrieval is often due to the one-size-fits-all approach: One type of warehouse is used for all items, regardless of volume or frequency of use. What characterizes an advanced factory is not that it uses a particular type of warehouse but that it uses different types to meet different needs.

The most common type of warehouse is the single-deep pallet rack shown in Figure 13.3. It provides random access to slots, meaning that every pallet can be retrieved without moving anything out of the way first. All levels are accessible through forklifts, narrow-aisle trucks, or order pickers, but the bottom level is more easily accessible, using pallet jacks to retrieve whole pallets or carts to retrieve boxes from pallets.



Figure 13.3 Single-deep pallet racks

Many plants use this concept and nothing else. Within this structure, however, there are still options on how items are stored. Items can be permanent residents with dedicated slots, or "hotel guests," with slots allocated to pallets as they arrive, which is also called "floating" or "chaotic" inventory positions. Again, many plants either have dedicated slots for all items, or dynamically allocated slots for all items.

Figure 13.4 shows the storage and retrieval traffic generated by the use of dynamically allocated slots for a heavily used item in an off-site warehouse. The item arrives in sea containers and goes to the production plant in full trailers. In between, the pallets are spread throughout the warehouse only to be collected again. Figure 13.5 shows the simplification of the flows for this item brought about simply by dedicating a zone to it. Incidentally, eliminating the traffic for this item in the rest of the warehouse also makes storage and retrieval of other items easier.

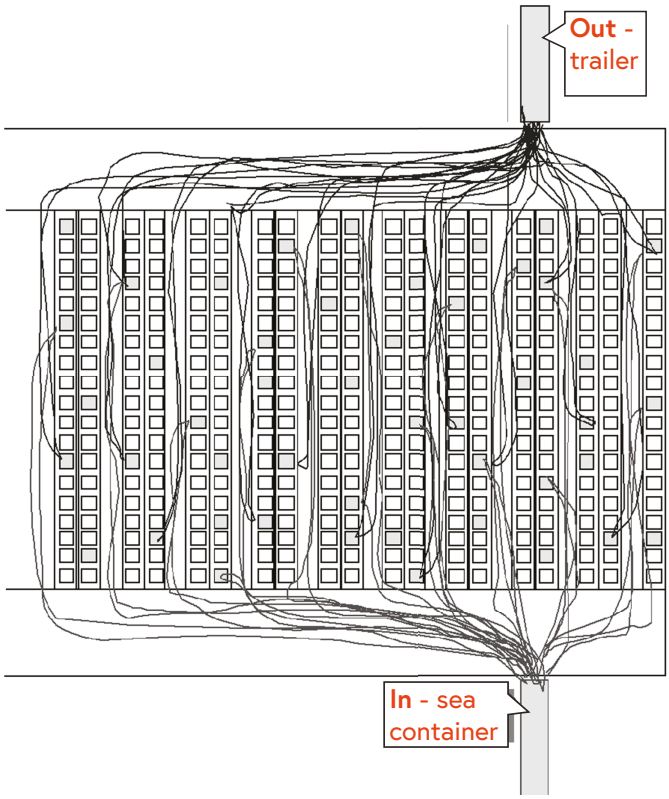


Figure 13.4 Dynamic allocation of slots to a high-volume item

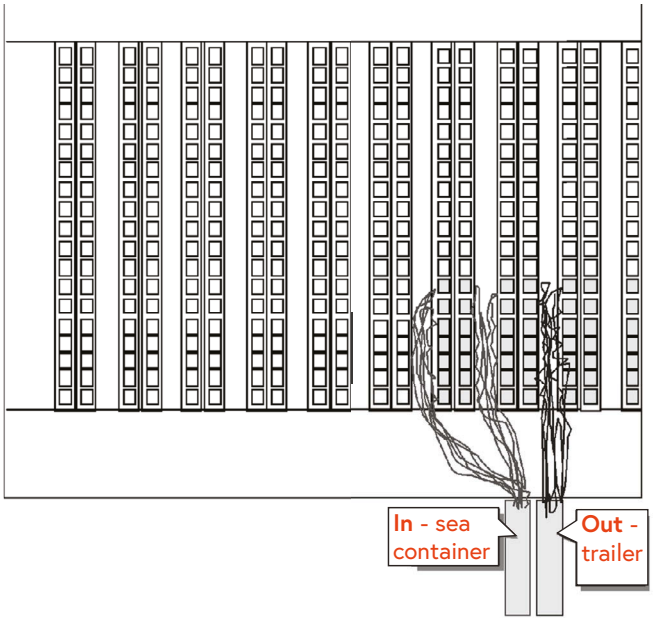


Figure 13.5 Dedicated zone for high-volume item

The approach in Figure 13.5, of course, is only the first step of improvement. Since the item arrives in sea container loads and leaves in trailer loads, there is no real need to put away the pallets on racks. By clearing up space for it, it should be possible to cross dock the pallets directly from sea container to trailer. The final step would be to eliminate the item's visit to the warehouse altogether and find a way to receive it at the production plant directly from sea containers.

This warehouse was designed to be flow-through, with materials coming in on one side and leaving on the opposite side. Figure 13.5 violates this principle, by having the outgoing trailer next to the incoming sea container. This, however, is a small price to pay to eliminate the unnecessary traffic, and, if necessary, the outgoing traffic can still be at the opposite side.

If an item is used frequently but in small quantities, the concept of having dedicated space remains valid, albeit on a small scale. The dedicated space may be no more than a slot holding one bin on a plain shelf. Where the concept of dedicated space is not valid, however, is for infrequently used items. Figure 13.6 shows a finished goods warehouse layout designed to accommodate both high-volume items shipped daily and infrequently needed items.

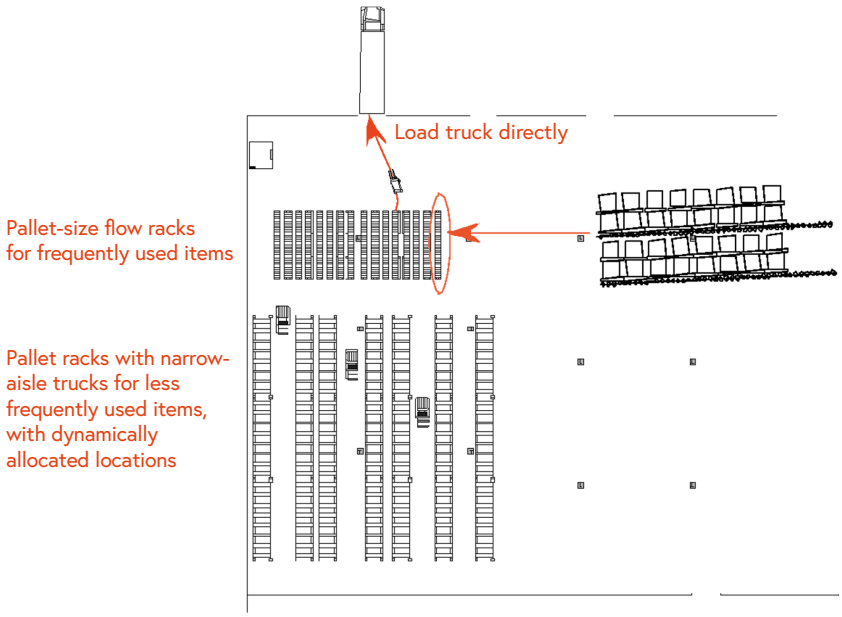


Figure 13.6 Warehouse for both frequently and infrequently used items

The motivations for using only one storage method when the combination of several would work better vary. A common one is familiarity. In his or her professional life, the warehouse manager has only been exposed to one approach and is uncomfortable with any other. Another one is the limitations of the Warehouse Management System (WMS) in place, only supporting one method of storage and retrieval. This is a reversal of normal roles, with the people as servants of the computer system rather than the other way around. Lack of support by the WMS should not be allowed to prevent the use of different methods, with the information it generates handled manually until the WMS is enhanced or replaced.

13.2.2. Volume occupied by finished goods versus raw materials

Finished goods rarely occupy the same amount of space as the materials they came from. In activities like extraction, refinement, or smelting the finished goods are smaller than the raw materials; in most processes, however, it is the opposite, with finished goods containing more air than the materials, and protected by bulky packaging.

All the materials that come into a factory eventually leave it, either as part of finished good or as scrap. In some activities, scrap is limited to disposable packaging materials and irreparable defectives; in others, a large proportion of the materials themselves are scrapped as part of normal operations. This includes not only extraction processes, but also machining. In conventional aerospace machining, for example, coffin-shaped forgings are hollowed out to 10% of their original weight while the remaining 90% returns to the forging shop as chips to be recycled, as shown in Figure 13.7. Note that this ratio can be reversed with additive manufacturing (AM) as discussed in Chapter 4.

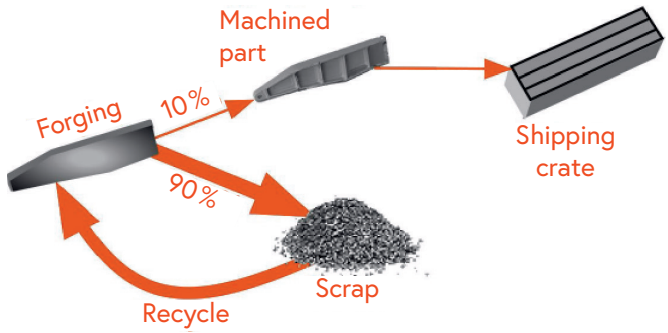


Figure 13.7 Conventionally machined aerospace part

The finished part takes up as much space as the original forging, and more once crated. The scrap metal in chip form also contains more air and takes up more space than it did when part of the incoming forging. In addition, the finished part is much more sensitive to handling damage than the original forging, on which a thin oxidized layer didn't matter since it was going to be machined away.

Figure 13.8 shows another product taking up more and more space as it moves from raw materials to finished goods. Not only does a finished, packaged frying pan take up 20 times more space than the aluminum disk, the coating, the handle, the rivets, and the carton, but then even the package must be protected from handling damage.

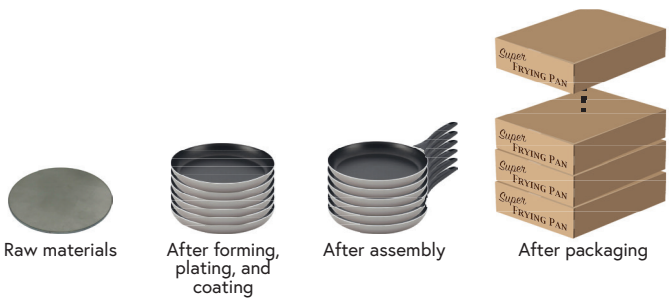
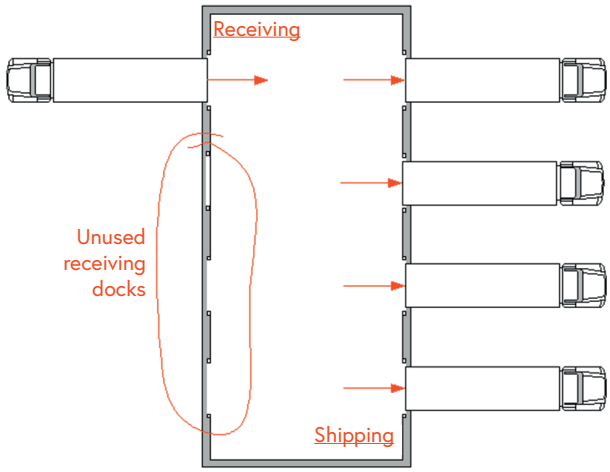


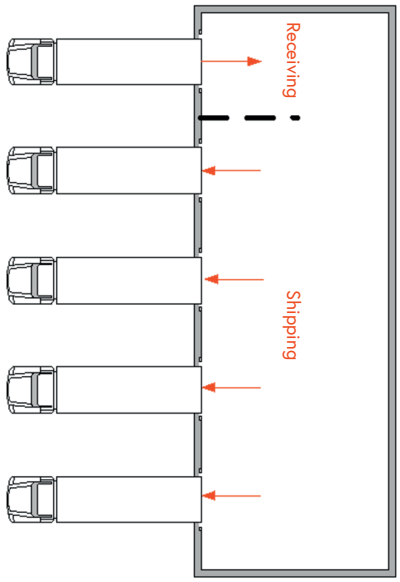
Figure 13.8 Bulk of finished goods versus raw materials

Where it occurs, this expansion in volume has consequences:

- Warehouse space that can hold a month of raw materials may hold only a day of finished goods. Forgetting this may lead to wrong conclusions about existing warehouse sizes, or to mistakes in warehouse design.
- Shipping from the factory may involve many more trucks than receiving. Docks used for shipping may be much busier than those used for receiving, or there may be many more of them. As shown in Figure 13.9, a flow-through building with equal dock capacity for receiving and shipping may not be as well-suited as one where both activities are collocated on the same side.
- These considerations have an impact on plant location. If there are 20 trucks going to customers for each truck delivering materials, then you want to minimize the transportation of finished goods by locating plants near customers. If, on the other hand, the plant refines a raw material and 90% of the output bulk is scrap, then the volume shipped to customers is small and it makes sense to locate the plant near the source of raw materials.



a. Flow-through building with dock number imbalance



b. U-shaped flow within building

Figure 13.9 Balancing shipping and receiving dock capacity

13.2.3. **Warehouse visibility**

In the management of a warehouse, visibility is essential. Surprisingly, it is often obvious when walking through plant warehouse areas that simple visual techniques are often left unattended. Warehouses need good visibility for effective, efficient, and safe operations. In factories, however, we often encounter warehouses with the following problems:

- **No addresses.** The location addressing scheme is not used and the racks are uncharted space.
- **Inconsistent location addresses.** In some aisles, the ground floor is level 0; in others, level 1. In some aisles, level numbers go up from bottom to top; in others, down from top to bottom.
- **Inadequate displays of location addresses.** People must search for location labels written in small or complex letters.

Again, no solution fits all. Steel bars need other storage systems than screws; paint barrels other systems than computer parts. There are, however, some commonalities that are generally smart:

- **Fit for purpose.** Fit storage systems to the parts, not the other way around. It depends on volumes, weight, turnover time, and other product and process characteristics.
- **Consistency.** A labeling regime should be consistently followed in a warehouse. There is no global standard, but there should be a local one.
- **Completeness.** Every location should have a an address. No product should be allowed where there is no address.
- **Visibility.** Print visual labels with large bold characters for human readability, and include bar-codes or RFID tags for automatic identification. Use arrows to indicate where to look.
- **Logical.** Label racks from the ground up. The lowest level should be named "01," the second level "02," and so on. Use a zero in all numbers less than 10.

Location labeling on the warehouse floor

Most warehouses are large rooms supported by a grid of columns, and simply affixing grid location labels on all sides of these columns makes it much easier to communicate with forklift drivers. Hence, even before any inventory system is implemented, the warehouse can be separated into a floor grid. The columns can be labeled accordingly. An example is shown in Figure 13.10.

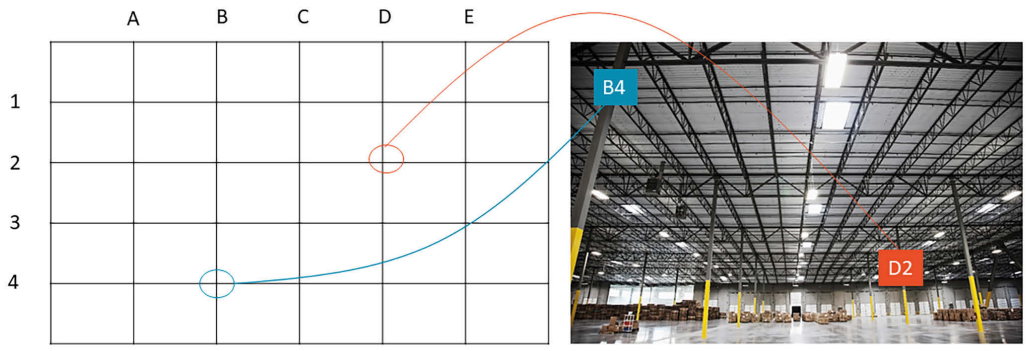


Figure 13.10 Column grid labeling

In warehouse areas with labeled racks, column labeling used together with rack labeling can create confusion. When racks are labeled, columns need no or different labeling notation.

Location labeling on racks and rack identification

A standard, effective, and self-explanatory method for identifying locations in pallet racks is by 4 coordinates for aisle, rack, level, and bin. Figure 13.11 illustrates the following definitions for common terminology in warehouse operations:

- **Aisle**: The walk way between two shelf racks. Each aisle can be marked with one identification – same for left/right racks – or with two, one for each left/right rack.
- **Rack**: A rack is a storage system that stores materials in horizontal rows with multiple levels.
- **Bay**, also known as rack unit): A bay is the basic storage unit in a rack and defines the pace between two vertical rack columns.
- **Levels**: A level refers to the vertical storage places in a rack.
- **Bin**: A bin is a a defined location that contains one or more spaces or slots for products. It is marked with a location identifier, usually a barcode. It is the most specific warehouse location.



Figure 13.11 Aisle, rack, bay, levels, and bin

Figure 13.12 shows how labeling is done effectively in an IKEA warehouse store. Note that IKEA splits the aisle in two rack numbers, in this case 12 and 14, which follows the logic of housing numbers in a street where all even numbers are on one side of the main "street" and all odd numbers on the other. Note also how the color and the shape of the aisle signs enhances their visibility. At IKEA, the two lowest shelf levels are picking zones for customers or staff, and the higher levels are storage space.



Figure 13.12 Aisle markings at IKEA (credit: IKEA)

IKEA uses numerical values to indicate aisles, Other warehouses use alphabetical values, A to Z, or AA to ZZ, and others again use alphanumerical values (as in Figure 13.11). As discussed, some use the same number for both sides of the aisle, while others continue the counting for each rack across an aisle. Also for bay and levels, companies choose different nomenclature. There is no global standard.

Slot separation

Marking a slot identification on a shelf is not always sufficient to ensure that the intended items remain aligned with it. As shown in Figure 13.13, while there is little risk of pallets becoming misaligned within a 3-pallet rack section, this can easily happen on a plain shelf holding small bags or bin quantities of many different items, but it can be prevented by materializing slots through separators.



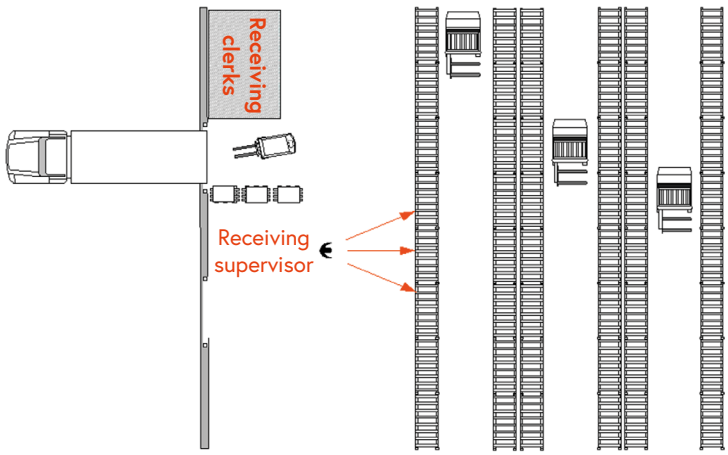
No slot separation: items can shift out of alignment with labels

With slot separators: item are kept in place

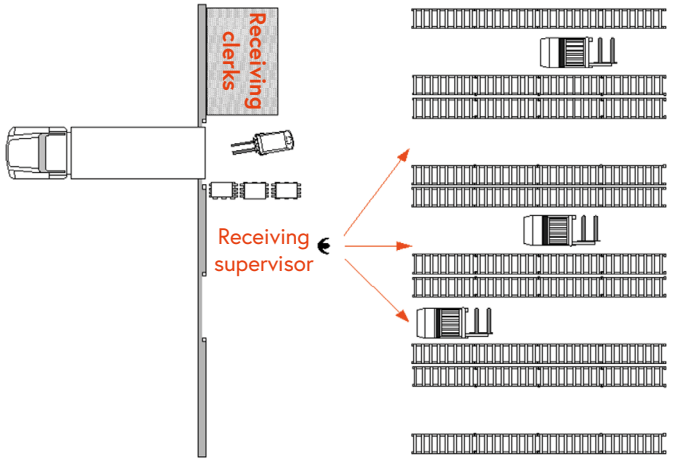
Figure 13.13 Shelves with and without slot separators

Rack orientation

Multi-level racks block the view. If single-deep pallet racks are oriented as in Figure 13.14a, then not only are the distances traveled to put away pallets longer than in Figure 13.14b, but the inside of the warehouse is hidden from view to all receiving personnel.



a. Rack orientation hides the warehouse



b. Rack orientation makes the warehouse visible

Figure 13.14 Rack orientation for warehouse visibility

Dock identification

Some warehouses erroneously use "now-you-see-it-now-you-don't" dock numbers. Dock numbers are mostly useful when the docks are open, as a means of helping forklift drivers avoid mistakes, yet many docks are labeled as in Figure 13.15, with the number hidden by the open door. Numbers printed on a roll-up door itself are also common. Numbers should be written so they can be seen at all times, such as for example on the side of the dock.

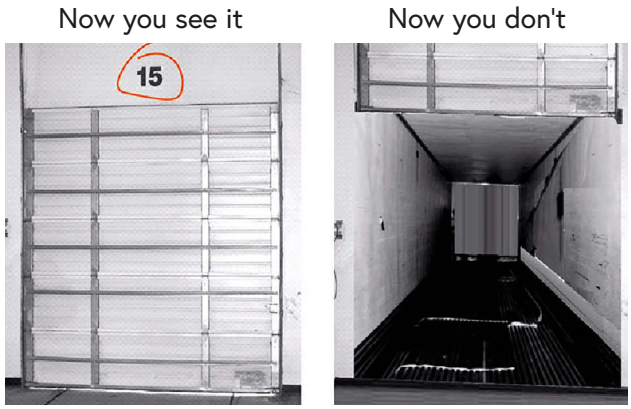


Figure 13.15 Dock numbers hidden by door

Zone identification

Zone identification is particularly important when floor storage or block stacking is used. *Floor storage* is as the name suggest storage of pallets or other stock units directly on the warehouse floor; it does not require any type of storage rack. In *block stacking*, loaded pallets are placed directly on the floor and built up in stacks to a maximum stable storage height (Figure 13.16).



Figure 13.16 Floor storage and block stacking, but in what zone?

Zones are often separated by painted lines on the floor, which is needed to clearly mark off where materials can be stored and allow effective walk and transportation routes. Maybe because someone is already doing the marking, zone name labels are also often painted or taped directly on the floor. That is, however, ineffective because the name of the zone is then only visible from up close, and only for a few weeks until it is erased or torn away by traffic, or hidden by pallets and goods that stand on it.

A common solution then it to hang a shingle from the rafters above the zone itself, but it is not visible from every direction, and is also subject to sway in the wind of the fan. A much less common but much more effective solution is to use triangular signs and mount them rigidly either on the rafters or above the racks. It stays in place and is visible from all directions. Again, IKEA serves as a good example of

how this can be done, as shown in Figure 13.17. Notice the triangle form of the signs, and the use of red color and large font size.



Figure 13.17 Good zone identification with triangular rigid signs at a IKEA warehouse

Stop and Think 13.1
Why do you think IKEA repeatedly features as a best practice company for warehouse management?

Serpentine and standard rack labeling

Consider the case when the pick list is in order by location. Two "best practice" labeling schemes for racks has been suggested by industry experts. It separates the *serpentine* and the *standard* method. In the standard method (see Figure 13.18), parts are stored as if pickers always start from the same ends of the aisles. In the serpentine system (see Figure 13.19), the picker will follow U-patterns in aisles to pick products for several orders. In the standard system, the picker usually has to double back to the beginning of the aisle.

Many experts argue in favor of the serpentine system. Why? Because when provided with a pick list *that is in order by location*, pickers can weave, like a snake, up and down the aisle and collect products for several orders without having to double back or skip around their pick list. It means a more organized and efficient route for pickers to follow.

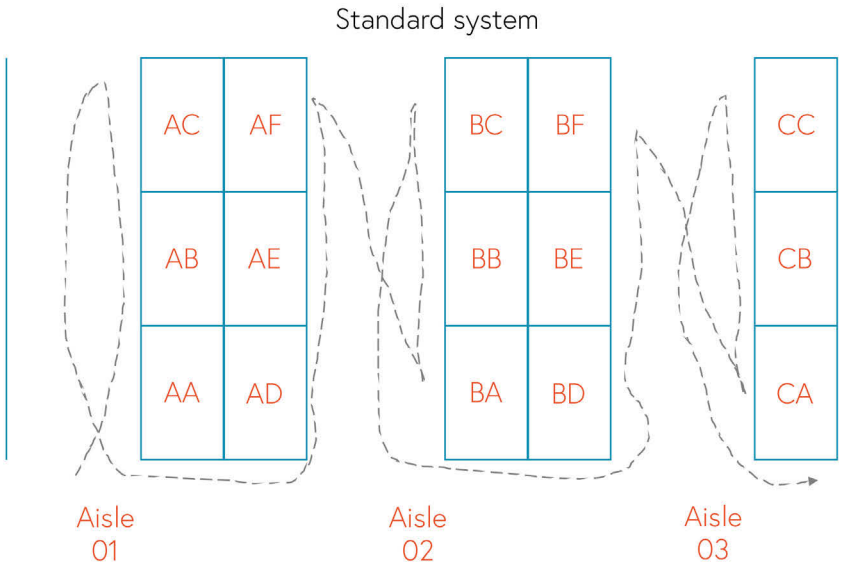


Figure 13.18 Standard labeling system

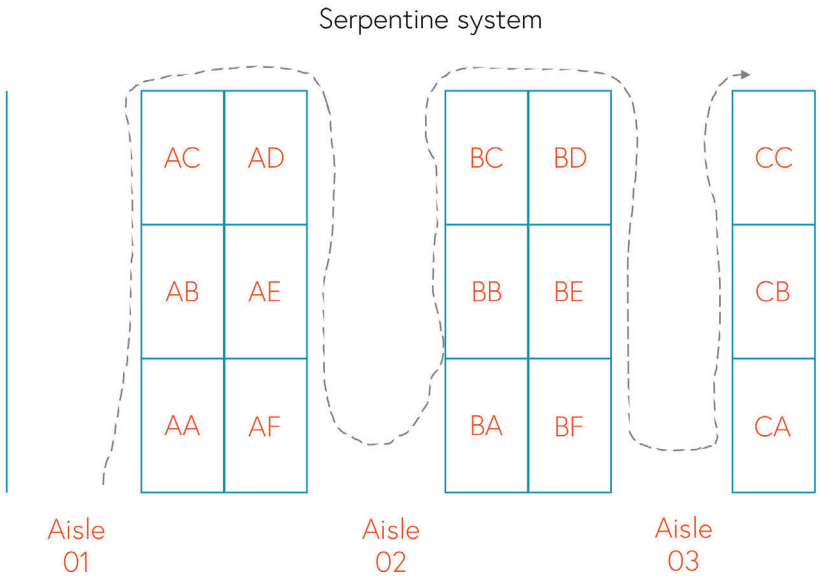


Figure 13.19 Serpentine labeling system

The standard and serpentine systems only apply when parts have fixed locations. They do not apply when a warehouse uses random placement and lets the WMS keep track of locations. In that case, optimization algorithms are used to calculate the shortest picking route.

13.2.4. **Slot allocation**

Any warehouse needs to decide what to store where. This decision can be made with help of simple rules or complex algorithms. Many academic articles are written about this topic and the latest trend is to use artificial intelligence to dynamically place incoming goods and change picking routes depending on the near-real-time situation. Such solutions can help improve productivity in e-commerce, retail, and distribution centers. In most cases in manufacturing, however, these solutions may be over-engineering, only serving to confuse and disempower operators. There is a value to logistics operators knowing where to find the material. When products and assembly sequences are fairly stable, a near optimal solution can be much simpler and work well. In almost every case, a good WMS will be of great assistance.

From simple to more advanced, 3 types of shelving systems are:

- Fixed slot allocation by item number – such as the library system
- Semi-permanent slot allocation by item number
- Dynamic slot allocation with manual card-based tracking
- Dynamic slot allocation with WMS

The library system

Some plants store items left to right and top to bottom in order of their item numbers, like books in a library as in Figure 13.20. Since it is easy to understand and provides a systematic way to find any item, it might seem a good idea. In general, however, it is not. Since frequency of use or destinations for any item are subject to change, they cannot be encoded in item numbers.



Figure 13.20 Shelving by item number as in a library is intutitve but not efficient for a factory

Organizing by item number harms operations in several ways:

1. The most frequently used items may be in hard-to-reach places.
2. Items that go into the same product may be located far from each another.
3. Introducing a new item requires shifting all the items that are after it in the number sequence.

Dynamic slot allocation

Instead, we should use a classification criteria, as a guide in slotting items:

1. **Frequency of use.** Frequently used items have easily accessible, dedicated locations; infrequently used items, dynamically allocated slots.
2. **Volume.** High-volume frequently used items may use flow racks; low-volume frequently used items, dedicated slots in racks or shelves.
3. **Destination.** Items from reliable suppliers used only in one area of the production floor should be collocated, as should items shipped exclusively to one customer.
4. **Source.** Items from a supplier with quality or delivery problems may be collocated for easier monitoring. This does not apply only to inbound materials, but also to internally made products about which customers have issued quality or delivery problem reports.

Depending on such circumstances as the size of the warehouse, the commonality of the items, or quality and delivery performance, these different criteria may be applied in different sequences. In addition, physical characteristics may also cause some items to be treated differently. Liquids or powders used frequently and in high volume may have dedicated silos and automatic transportation to points of use.

Flow racks for high-volume, frequently used items can be arranged by order of their item numbers, but it will not make much difference. If there are 5 to 10 of them, stored or retrieved hourly, materials handlers promptly memorize where they are. Ordering by item number is useful for the 50 or 100 items that have dedicated slots in a pallet rack, but it is not usable at all for the thousands of items that use dynamically allocated slots. They must be assigned whatever slot is available when they arrive, and the handlers must register these "hotel guests" for later retrieval.

In the example of Figure 13.21, IKEA directs consumers to aisles from a list of item families in alphabetical order. The assignments are semi-permanent, in that they may change once a season but not everyday. The map is needed because customers cannot be expected to know this information.

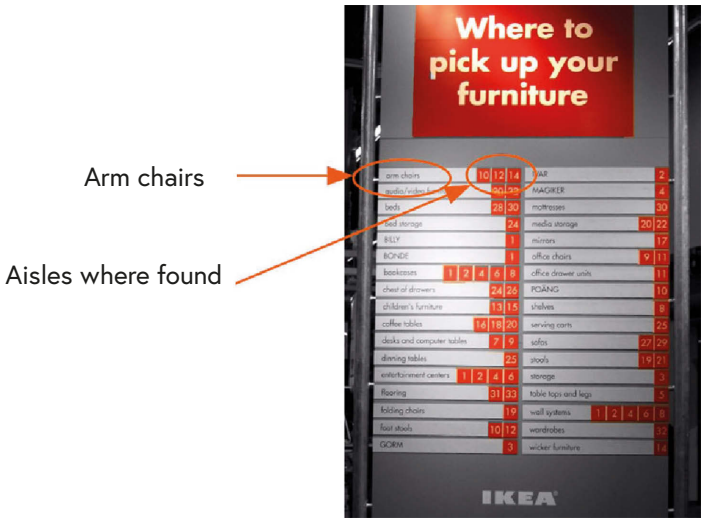


Figure 13.21 IKEA's item location guide

Dynamic slot allocation with manual card-based tracking

Figure 13.22 shows an approach for manual tracking of dynamically allocated slots. Each slot has a "key" attached to it when it is empty. When loading a pallet into the slot, the materials handler takes away the key, detaches a duplicate identification tag from the pallet, and places both in the corresponding position on the map at the front of the aisle. Conversely, when retrieving the pallet, the handler takes along the key and the tag.

This approach avoids having to write anything down by hand, but it is still quite vulnerable to employees being distracted or not properly trained. Also, while it allows you to find easily what is in location 3b-4-2, this is not the most relevant question. Instead, what handlers most frequently need to know is "Where do I find item 52103?" With only the information in Figure 13.22, the handler has to visually scan the maps of all the aisles. With 10 aisles, this would be prohibitively time consuming. If, however, zoning is used to reduce the allowable storage areas for an item to 1 or 2 aisles, it may be manageable. Alternatively, a clerk could periodically scan the maps and maintain a catalog of item locations.

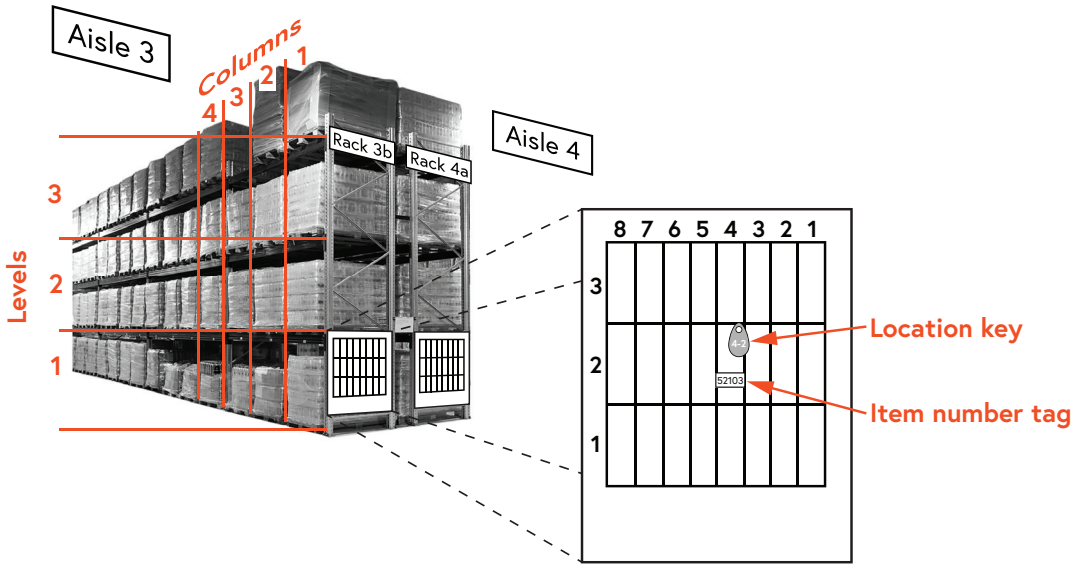


Figure 13.22 Manual mapping of a pallet rack

Dynamic allocation with Warehouse Management System

The labor required and the vulnerability of the manual systems for tracking dynamically allocated slots may explain why computer-based WMS, as shown in Figure 13.23, are possibly the most successful application of information technology to manufacturing. Unlike ERP or MES, WMS do not generate much controversy. It is the consensus of materials managers that these systems pay for themselves through reductions in data collection labor, higher accuracy, faster retrieval, support for cycle counting, and data mining.

A WMS is a software application that helps manage and control the operations in a warehouse or distribution center. A WMS guides inventory receiving, placement, and retrieval. It also helps optimize

picking and shipping of orders and can notify when inventory should be replenished, is slow-moving, or obsolete. A modern WMS offers apps for mobile devices and integrates with barcode and RFID scanning.

In many warehouses, barcode scanning has been in use for decades, for functions like matching serial numbers to customer purchase orders (PO), and the hardware and communication infrastructure to support a WMS is already in place. In the example shown here, the materials handler scans barcodes for both location and item number when loading a pallet. The WMS stores the pair of codes in an item-location cross-reference table, which can then be queried for the location of any item number. Barcode scanning is not perfect, and is being superseded as an auto-ID technology by radio-frequency identification (RFID).

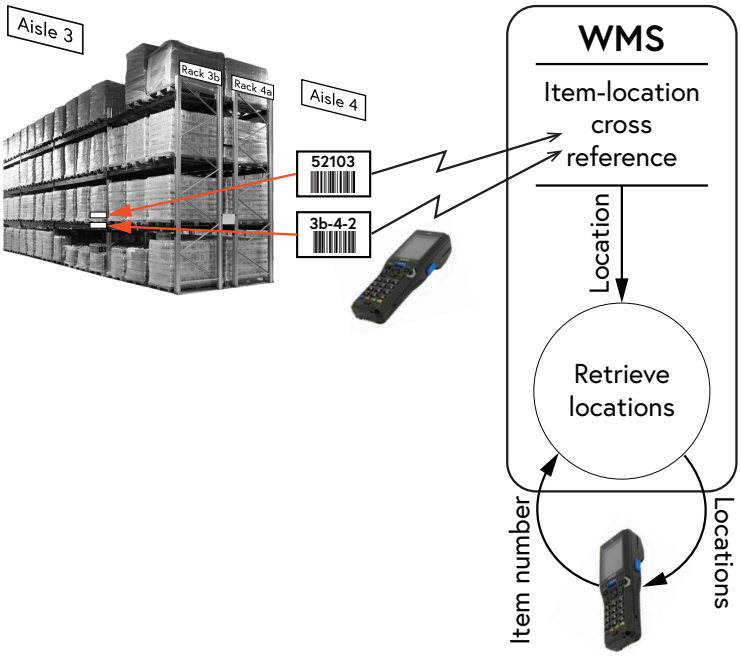


Figure 13.23 Mapping of a pallet rack with a WMS

A WMS can be a standalone application implemented locally or in the cloud, or part of a larger ERP system. ERP vendors often include WMS as a module in their system offerings. A WMS system can also cover extended capabilities such as labor management, yard management, kitting, pick-by-voice/light, and other activities.

Stop and Think 13.2
Why do you think WMSs, generally speaking, have more satisfied users than for example ERP and MES systems?

13.2.5. **Differentiated use of the vertical levels**

The multiple levels in a pallet rack can be given different roles. It is common to keep the first or two lowest shelf levels as picking zones and the higher levels as full storage units. The lowest zones are easily accessed by human workers who can pick goods without having to elevate with forklifts or scissor cranes. Figure 13.24 shows the concept in a distribution center for fast-moving consumer goods. The lowest level is reserved for pallets that are ready for picking parts of the pallet. Forklifts periodically come by to bring pallets down to the floor level and replenish the upper levels.



Figure 13.24 Specializing shelf use by level

At IKEA, for example, customers pick at the lowest 2 levels, and IKEA utilizes the rest of the higher shelf space for warehousing. Sometimes the second lowest shelf is used to display the assembled products. Figure 13.25 shows a similar approach for a warehouse where materials arrive in pallets of bins and leave in individual bins, mix-loaded on a train of tow carts.

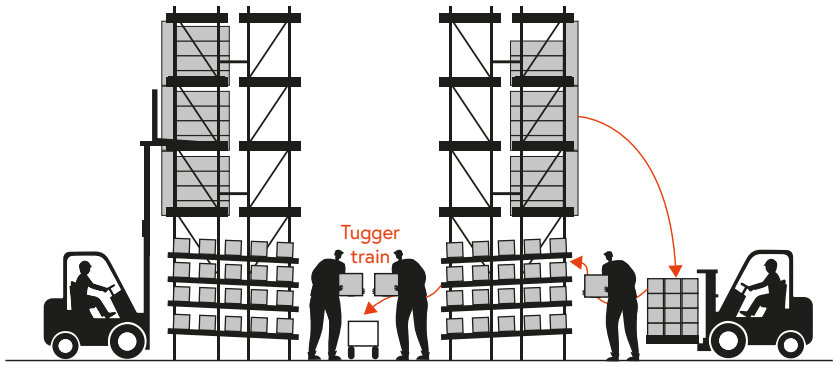


Figure 13.25 Rack/flow rack combination

13.3. Manual versus automated storage and retrieval

Storage and retrieval systems come in a variety of shapes and modes of operation, and, there again, tailoring the choices to the different needs results in a cheaper system that can perform better than one-size-fits-all. Whether it does perform better, of course, depends on how it is used and managed.

13.3.1. Manual storage and retrieval systems

Figure 13.26 shows a few alternative rack designs for manual pallet handling.

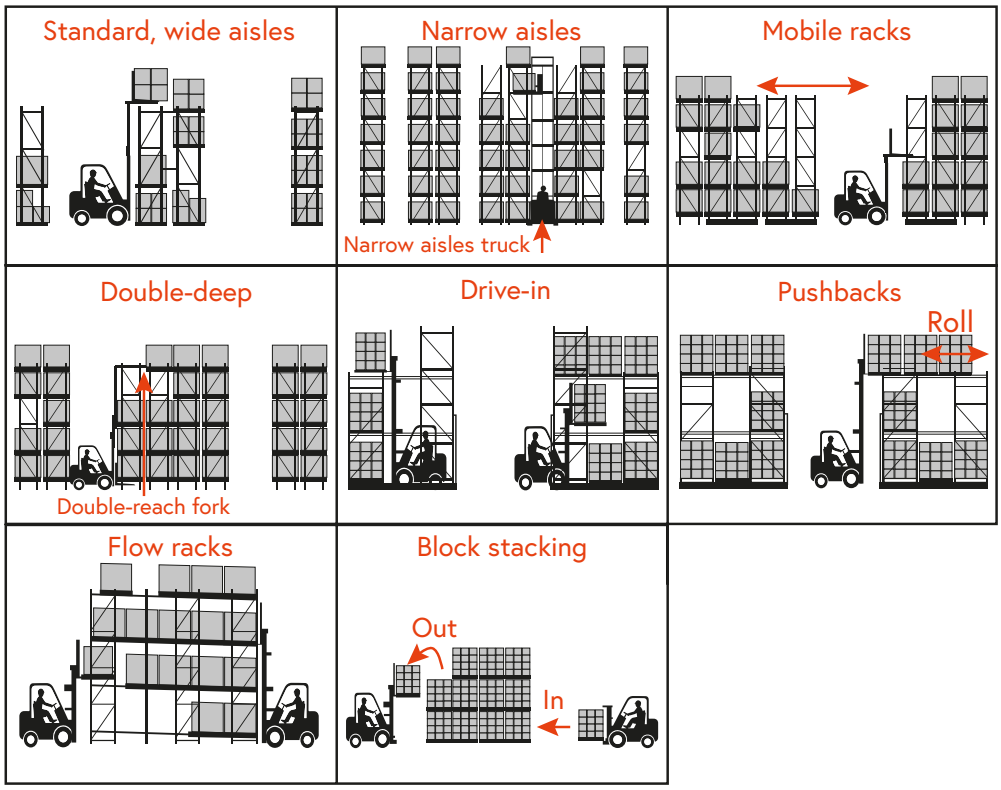


Figure 13.26 Pallet storage options with manual handling

Narrow aisles or *mobile racks* are variations on *single-deep* pallet racks that support a higher storage density, at the cost of requiring special equipment. The investment in equipment for narrow aisles may be easier to justify if it allows the handler to ride up into the higher levels and pick individual boxes from pallets.

Double-deep, drive-in, and *pushback* racks also increase storage density, require special equipment, and allocate storage in multiple pallet units with last-in-first-out (LIFO) retrieval. Storage and retrieval requires less labor than single-deep racks.

Flow racks are for items with dedicated storage space. They support first-in-first-out (FIFO) retrieval at the pallet level, and require the least labor of all the approaches listed. *Block stacking* stack pallets without racking infrastructure.

Block stacking

Besides single-deep racks and flow racks, block stacking is the only approach in Figure 13.26 to be widely used in manufacturing. The main attraction of block stacking is that it requires no investment in racking and relies on the same forklifts the plant uses for other purposes. It requires pallets to be stackable, which means that the container walls must be sturdy enough to bear the weight of several pallets. For a labor point of view, it is the polar opposite of flow racks, which rely on gravity to move pallets. Block stacking is the most labor-intensive method, because nothing moves unless a forklift moves it.

By design, block stacking almost always ends up as last-in-first-out (LIFO). However, as shown in Figure 13.27, stacks of pallets can be arranged in first-in-first-out (FIFO) lanes as cars queuing in multiple lines for a ferry.

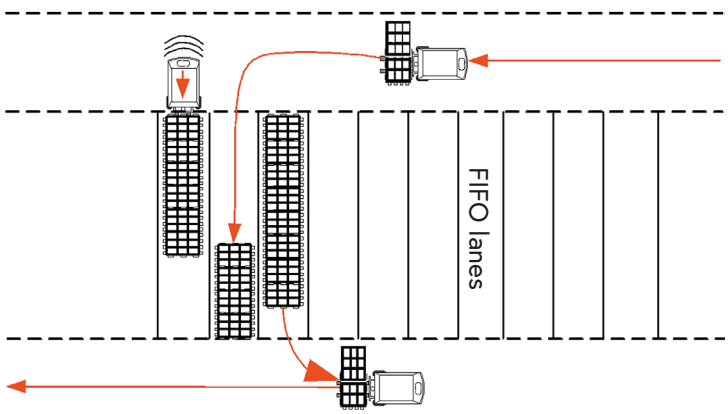


Figure 13.27 Storage and retrieval with block stacking in FIFO lanes

Since one entire lane is used for a single item, it supports considerably fewer items than systems that rely on racking to store different items at different levels over the same real estate. These characteristics make block stacking appropriate for high-volume-low-mix applications, with items shipped by the truckload every day. Because of its low setup cost, it tends to be used when a plant first starts up, even where it does not fit long-term requirements.

13.3.2. Automated Storage and Retrieval Systems (AS/RS)

None of the approaches in Figure 13.26 involves automation. All movements are power-assisted, but all actions are taken by people. With *Automated Storage and Retrieval Systems* (AS/RS), as shown in Figure 13.28, operator action consists of either (1) entering materials into a loading station and the loading transaction on the control system, or (2) requesting materials from the control system and unloading it on delivery. The selection of a slot and the movements of the materials to and from that slot are controlled and executed without human intervention, and faster than in the manual systems.

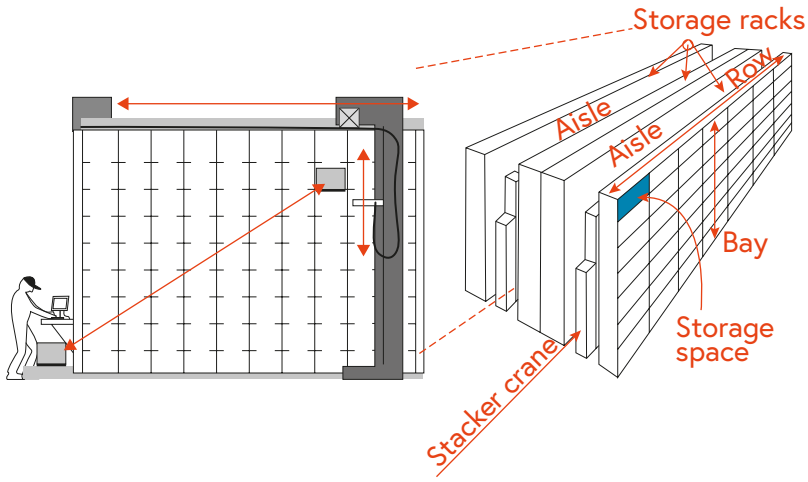


Figure 13.28 Working with an AS/RS

AS/RS can be economically feasible when there are very high volumes, storage density is important because of space constraints, and inventory accuracy is critical. The benefits of an AS/RS are reduced labor for transporting items into and out of inventory, more accurate tracking of inventory, and space savings. Figure 13.29 shows the scale and complexity of a large AS/RS.



Figure 13.29 Automated Storage and Retrieval System

Like AGVs, AS/RSs have existed since at least the 1980s, yet they are far from common in manufacturing warehouses, and, where found, their users often express buyer's remorse. However, it is one decision to acquire an AS/RS when you don't have one, and another to dispose of one that is already bought, paid for, and in use.

Much of the dissatisfaction users experience with AS/RSs can be traced back to the acquisition procedure. When ramping up a new plant or a new production line, a rational approach could be to start

with block stacking, which requires no investment, migrate to flow racks and pallet racks with manual operation as you outgrow block stacking, and finally, once operations are well established and the benefits of automation can be assessed, invest in AS/RS technology as needed. The budgeting process of many companies, however, causes plant designers to include AS/RSs in the initial capital acquisition request. As a $5M line item in a $300M project, it will be more easily approved than as a standalone $5M project two years after plant startup, when management is more focused on having the plant make money than on spending to set it up.

As a consequence, AS/RSs are designed before the organization has an opportunity to understand and specify its requirements, and this often results in discovering a mismatch after the fact, as in the following two examples:

1. **Control system.** One factory bought an AS/RS with a control system that could only store and retrieve full pallets, while its operations required picking individual boxes from pallets. The factory had to set up a manual warehouse for partial pallets next to the AS/RS, which largely erased the benefits expected from it.
2. **System sizing.** A machine shop sized an AS/RS for cutting tools based on the assumption that all tools would be stored away after each use in a machining center. After the factory started up, users found it more convenient to cache frequently used tool sets outside of the AS/RS than to separate and store each tool individually, leaving the AS/RS 3/4 empty.

Besides very high investment costs, the biggest challenge with an AS/RS is when it doesn't work. Many manufacturers have experienced the painful retrieval of needed parts from an AS/RS that has unplanned downtime. Imagine if the transport system in the AS/RS shown in Figure 13.29 stops working. In addition, AS/RSs have the following limitations:

- **Lack of visibility.** Inventory in an AS/RS is only visible through its control system, and AS/RSs have been denounced as black holes where unnecessary inventory disappears from view.
- **Lack of flexibility.** The operating policies that can be used with an AS/RS are limited to what its control software will support. Manufacturing organizations often find that it does not meet their needs and don't have the resources to modify it.
- **Impact on manual storage and retrieval operations.** An AS/RS can spoil its users into not using the normal visible controls in the manual stores they still need to maintain.
- **Focus of attention.** In plants that have an AS/RS, its effective use becomes the focus of all debates about materials management. That it is more expensive than all other materials handling resources together makes it difficult for managers and engineers to think about any other issue.

13.3.3. Carousels and storage lifts

Carousels are simpler and cheaper than an AS/RS, primarily because they move along only one axis, as shown in Figure 13.30 in both a horizontal and a vertical version. The vertical carousel is a more complex machine but uses less floor space. The movement of a carrousel may be controlled by the operator or by a control system that knows which items are stored where.

With a vertical carousel, also known as *paternoster*, all storage locations are brought to the operating station by a cyclical circulation. Modern systems instead use *storage lifts,* which collect and bring only the part needed to the operator. A storage lift is a AS/RS for small parts.

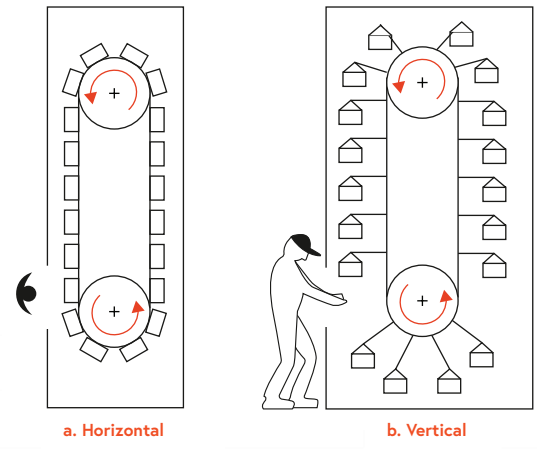


Figure 13.30 Horizontal and vertical carrousel concepts

In everyday life, carousels are commonly found in dry-cleaning stores – where they allow one attendant to retrieve a customer's garments while remaining by the cash register – and in airport baggage claim areas — where they allow luggage loaded at a single point to be unloaded by passengers distributed along the entire perimeter (see Figure 13.31). An everyday example of a storage lift is the vending machine.



Figure 13.31 A carrousel in everyday life: Single put-away point below ground and distribution of luggage along the entire perimeter of the carousel

In most manufacturing situations the case for carrousels is not very compelling. In most storage and retrieval operations, the sequential access provided by a carrousel is a serious drawback, since, in order to reach a target item, it may have to scan through many others, leaving the operator waiting almost as long as it would have taken him or her to retrieve the item by other means. Modern storage lifts systems provide an improvement but do not remove obstacles like bottleneck issues and lack of visibility.

Vertical and horizontal carrousels are sometimes used as line-side presentation devices to prevent picking errors on an assembly line. But, even for this purpose, it has limitations. If the line assembles a batch of different products of a fixed sequence, then bins can be placed in the carousel in a matching sequence, so that the carousel only has to move by one slot to set up for the next product. If, on the other hand, the line builds single units in a leveled sequence that is recalculated every shift, then the carousel would have to switch bins with every unit, and the bin for the next unit may be arbitrarily far from the current one.

Storage lifts have some advantages. When the factory stores many small but valuable parts, an automated system can increase inventory accuracy – given that all parts are diligently registered into and out of the system. The greatest advantage is the possibility to store many parts at a small space, but with the drawback that one creates a monument that is difficult to move. These systems can also only be operated by accepted personnel and operator IDs can be tracked during storage and retrieval. While the advantages of picking from the system what one needs when one needs it is palpable, what is often forgotten is that parts also have to be put into the system. When there is only one handling hatch, an automated storage system risks becoming a bottleneck.

Stop and Think 13.3
Over-the-counter pharmacies sometimes use storage lifts. Why?

13.3.4. Other AS/RS systems

Other automation systems for warehouses include, among others:

- **AutoStore**, a large-scale AS/RS, with a cubic layout where robots put away and pick from the top of racks. This system is used in many retail warehouses. In manufacturing, AutoStore can be of relevance for high-value products where accuracy and security has top priority, or where space is extraordinary limited or expensive.
- **Amazon Robotics** (previously Kiva Systems) moves the entire racks to the pickers with small mobile AGVs. This system is used in several Amazon fulfilment centers.
- **Forked AGVs** are automated forklifts that use vision and navigation technologies to automatically drive and operate the forklift.



Figure 13.32 Modern warehouse automation, from left: AutoStore (Credit: AutoStore), Amazon Robotics's AGV (Credit: Amazon), and Forked AGVs (Credit: Linde)

Due to their investment costs, these systems are more relevant for large distribution centers of retailers than for manufacturers. In some special cases, however, the systems are worth evaluating.

13.3.5. Pick-by-voice and pick-to-light

Pick-by-voice, or voice-directed warehousing, was introduced in the 1990s and has become commonplace in large warehouses and distribution centers. With pick-by-voice, logistics operators get instructions via headphones where to go and what to pick. Workers confirm by voice where they are and what they pick. This system replaces manual paper-based or handheld computer-based picking. The great advantage with pick-by-voice is that the operator's eyes and hands are freed from the task of getting picking information, which increases speed, safety, and accuracy. Arguably, it makes the picking task easier, which can reduce training time.

Pick-to-light directs manual picking by lighting up bins. While not extensively covered in the literature, you find it in many factories, where it reduces training time and picking errors, while increasing picker productivity. Pick-to-light is a form of *usability engineering*. It's not automation or jidoka because it just prompts the operator. And it's not mistake-proofing because it doesn't physically prevent mistakes. It is more an intermediate technology, a stepping stone on the way to full picking automation. But is it? It can also be viewed as a move towards using technology to make work easier for people instead of automating. Figure 13.33 illustrates this point.



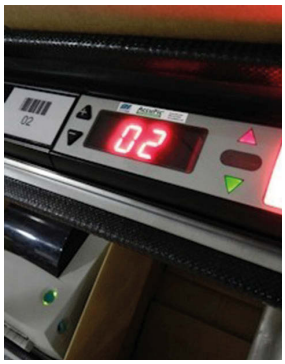
Figure 13.33 Picking support technologies from left: manual scanning, pick-by-voice (Credit: ACD Gruppe), and pick-by-light (Credit: KBS Industrieelektronik)

With pick-by-voice or pick-to-light, pickers no longer need to look for the right bins, and Materials Managers can change bin locations without confusing pickers. Because it supports manual picking rather than replace it, it is semi-automation. At a fraction of the cost of full picking automation, it has tangible benefits that make it popular in warehouses and manufacturing operations.

Manufacturers use pick-to-light primarily used for kitting in supermarkets and for lineside picking at assembly stations. On factory shop floors, "supermarkets" are small stores collocated with Assembly and owned by Production. The Materials organization replenishes the racks from the outside. Special production operators, often called "water spiders" pick kits from the supermarket and deliver parts for immediate use on the line. Assembly organizations have found pick-to-light to be a great help in picking accurately and promptly. To bypass the kitting step, you can apply pick-to-light directly on lineside shelves for operators to pick and assemble.

In both technologies, directing pickers with voice or lighting up bins are the easy parts. *Validating* picks in terms of items and quantities and reporting the pick to Production Control is more challenging. Some pick-to-light systems post the pick quantity on the bin and provide a button for the operator to press once finished. Others place a light curtain in front of each bin. Then the pickers break this curtain as they reach in and restore it when they pull their hands out, telling the system that they are done with the bin. This approach, however, still does not control the pick quantities. The light-curtain approach

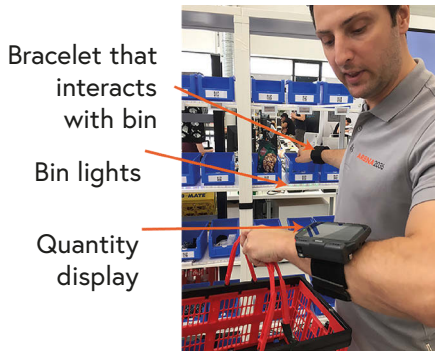
requires additional hardware at each bin that is only used when the bin is lit for picking or when the picker makes a mistake. If the picker wears this hardware instead, then you don't need to replicate it at each bin. The state of the art as of 2022 is a bracelet worn by the picker interacting with an RFID tag attached to the bin. These different validation methods are shown in Figure 13.34.



Manual pick confirmation
(Credit: Accupic)



Automatic pick confirmation
(Credit: Turck-Banner)



Wearable automatic pick confirmation
(Credit: ThingOS)

Figure 13.34 Manual pick confirmation method

Box 13.2 **Early pick-to-light ideas: Saridis' patent**

In 1973, Harry Saridis, Berkeley, CA, received a U.S. patent for a semi-automated retail store. In this store, customers would fill out order forms and feed them to a computer, which would light up bins for an attendant to pick from. A conveyor belt then delivered the filled order to the customer at the check stand. Figure 13.35 shows the patent.

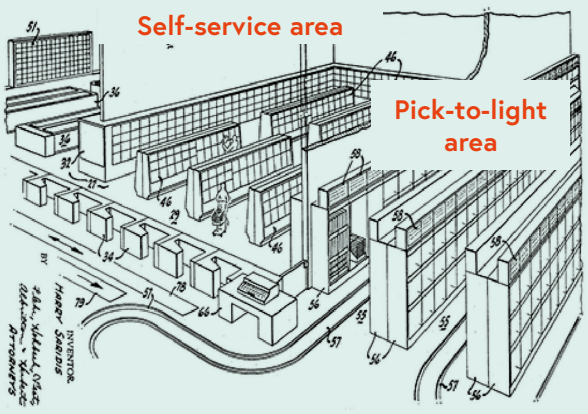


Figure 13.35 The Saridis retail concept

Saridis's application didn't catch on. He thought stores would use it to reduce shoplifting. In operations, however, it would have been a step backwards from the supermarket to earlier forms of retail in which customers did not have direct access to the goods. In converting to supermarkets, retailers bet that they could trust customers, who would enjoy picking their own fruits and vegetables and would supplement their shopping lists with impulse buys. And the stores saved picking labor at the same time. It was a win-win, and Saridis's concept was an expensive way to undo it.

13.4. Organizing the warehouse

The concepts and methods introduced in the next sections help making the warehouse effective and efficient.

13.4.1. Don't fill the warehouse full

Some managers are uncomfortable at the sight of a warehouse that is not full. They feel that it is wasting space. As a result, many factory warehouses are almost full all the time. But this is not in fact a desirable situation. The service warehouses provide is storage and retrieval, and you cannot store into a full warehouse any more than you can retrieve from an empty one. Let us assume the plant consumes one truckload per day of an item and has a dedicated location for this item in the warehouse, then, assuming 10% of a day's consumption in safety stock to protect against traffic delays, the amount of the item in storage will be as shown in Figure 13.36, with an average occupancy of 60%.

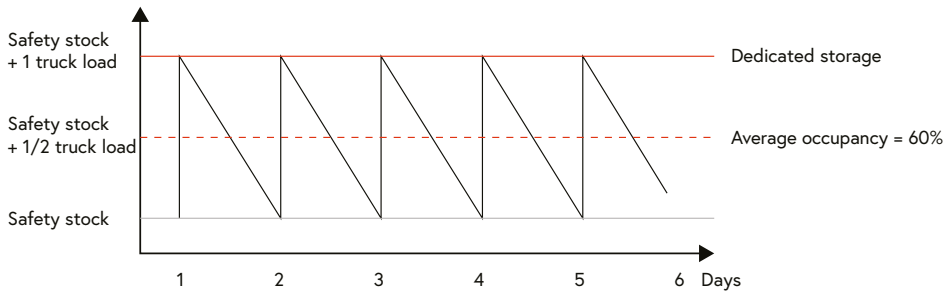


Figure 13.36 Storage of a heavily used item over time

In general, productivity of a warehouse suffers as occupancy exceeds 85% to 90%. Higher occupancy can be achieved when storage locations are dynamically allocated, because any space freed up by the retrieval of one item can immediately be used for the storage of another. There is no general formula for when storage and retrieval performance begins to degrade but specific designs can be simulated.

The phenomenon itself is commonly experienced when accessing computer hard disks or parking at a shopping mall, as follows:

1. To store files on a hard disk that is nearly full, computer operating systems fragment them into pieces that are small enough to fit in available slots, and puts the pieces back together when retrieving them. The time needed for these operations is perceptibly longer than storing and retrieving contiguous blocks of data from a nearly empty disk.
2. While parking does not involve fragmentation, it still requires hunting for a spot and returning to that spot afterwards. The best, most convenient spots go first, and only the most remote remain available when the parking lot is nearly full.

Assume a single-deep pallet rack warehouse, to which one item arrives by the truckload and leaves in multi-pallet loads while other items arrive and leave in varying quantities. Then both of the phenomena described above occur. When the warehouse is empty, truckloads can be stored in contiguous slots on the ground floor. When it is nearly full, they are fragmented into single pallet units wherever a slot is

available, and this will tend to be in the highest reaches of the racks that are farthest from the entrance, and multiple pallets need to be retrieved and assembled from different slots.

For illustration purposes, a simple formula can be obtained if we make the following simplifying assumptions:

1. As a result of ongoing storage and retrieval activity in a single-deep pallet rack, all slots have an equal probability of being occupied, independently of one another.
2. The warehouse is large enough so that the occupancy is significantly unchanged as a result of storing an additional truckload. For example, if the warehouse has 5,000 slots and is 90% full, then 4,500 slots are occupied. If we store another 20 pallets, then the occupancy rate rises by 20/5000 = 0.4%, which we can neglect.

The average number of slots that need to be visited in order to store a number of pallets is then equal to the average number of attempts needed to win a given number of times in a lottery where the odds are determined by the occupancy rate. This is given by the formula:

$$Number\ of\ slots = Number\ of\ pallets \times \left[1 + \frac{Occupancy\ rate}{(1 - Occupancy\ rate)} \right]$$

which is plotted in Figure 13.37.

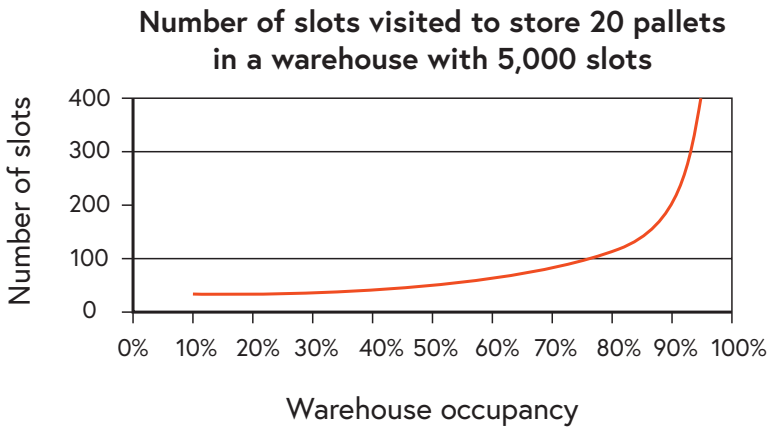


Figure 13.37 Simplified warehouse saturation model

Any real situation is of course more complex and not amenable to such a formula. For example, if a WMS provides directed putaway, then the forklift operator is told where to store each pallet and doesn't have to hunt for a slot. This formula also says nothing about retrieval of multiple pallets from a fragmented warehouse or about the fact that some of the open slots in a crowded warehouse are in hard-to-reach places. What we do know is that we can expect storage and retrieval performance to fall catastrophically as the warehouse saturates. *A priori*, we don't know where to draw the line, but we know we should expect this phenomenon as occupancy rises, and we can watch for its onset.

13.4.2. Track inventory transactions

Some plants have allowed production operators to withdraw parts from the materials warehouse without any tracking transactions, on the ground that record keeping was non-value added work and that these actions could be accounted for through back-flushing on finished goods. There are two problems with this approach:

1. Short of doing a complete physical count every day, the only way to know what is inside a warehouse is to record what goes in and what comes out. Not tracking withdrawals destroys inventory accuracy and prevents the generation of any actionable production schedule.
2. Just because a part is needed for one assembly operation does not mean that it is not also needed for another. The policy of letting production operators help themselves to materials resolves such conflicts by giving priority to whoever grabs the parts first, which isn't necessarily appropriate.

Where tracking transactions can be overkill is not in the warehouse but on the production line. If production is organized so that parts issued from the materials warehouse leave the line within hours, whether mounted into finished goods or scrapped as defective, there is no need to maintain a database of how many of which type of parts are within the production line. There may be a need, however, to attach the serial numbers of parts to the serial number of the product for traceability.

Withdrawals of parts from a warehouse should be based on the following principles:

1. **Allow only materials handlers to touch items in a warehouse.** Production operators or team leaders should not need or be allowed to fetch parts directly from the warehouse, and they shouldn't have the time to do it anyway. Delivering parts to production is the purpose of the materials organization, and it should fulfill it.
2. **Withdraw parts from a warehouse only based on some authorization signal from Production.** This is the pull principle applied to the warehouse. Materials handlers should not take the initiative of delivering unsolicited parts, and their supervisors should not instruct them to do so.
3. **Materials handlers always record item numbers and quantities when withdrawing parts from the warehouse.** Most warehouses have WMSs that reduce the transaction work to scanning a barcode. The move to the IIoT should completely automate the transaction and relieve the materials handlers of the need to enter data manually in any form.
4. **Make parts visible and easy to count.** For liquids or powders, use partially transparent containers to make fill level visible. For discrete parts, use dunnage to support counting.
5. **Maintain inventory accuracy by cycle counting.** The accuracy of inventory records should be enhanced by cycle counting, a practice in which materials handlers make physical counts of a few items every day on a rotating basis.

Full counts are disruptive, labor-intensive, and produce results that are not as accurate as they should be. They are snapshots of a warehouse, from which you deduce its content over time from in and out transactions but errors accumulate over time. Integrated in daily practice, cycle counting detects and corrects these errors without disrupting operations. The amount needed may drop over time as warehouse operations improve, but it is unlikely to vanish altogether.

Even companies that track placement and withdrawal and do cycle counting will have discrepancies in what the WMS says is in the warehouse and what is actually in the warehouse. Even with the best intentions and best systems, people make mistakes, goods expire, or other issues occur Even an experienced warehouse management company like IKEA have warehouse accuracy problems that hurt performance. Box 13.1 shows how IKEA tries to close the gap with new technologies such as drones.

Box 13.1 **Cycle counting with drones at IKEA**

Knowing what one has on inventory, and where to find it is essential to warehouse effectiveness and efficiency. The best way to ensure it is to continuously track materials into and out of warehouse locations, preferably in real-time and linked directly with the WMS. Nevertheless, discrepancies between real inventory and recorded inventory occurs. These problems may be lesser with automated inventory but are never solved completely. Put-away and picking and the recording of these activities are prone to human error. In addition, supplies may be wrong, and stored materials may be damaged or go out of date.

Inventory accuracy can be enhanced by *cycle counting*. Cycle counting is a method of checks and balances by which companies count physical inventory and check if it matches their inventory records. It is performed on a regular basis where different parts or a random sample of them are checked. During a year, all parts have been counted.

The problem with cycle counting is that it is a very manual activity and prone to errors itself. It is tedious and labor-intensive. Therefore, IKEA is among several retailers that have tried to automate cycle counting using drones. A breakthrough happened with the introduction of fully autonomous drones from drone vendor Verity based in Switzerland.[1] Verity's solution involves several drones that, during the night, scan rack positions that have had movement during the day, and every weekend scan all warehouse positions. Barcode scans and photos are automatically transferred to the WMS.



Figure 13.38 Cycle counting by Verity drones in an IKEA warehouse (Credit: Verity/IKEA)

Drones are exotic in manufacturing and therefore often seen in promotional "Industry 4.0" presentations and videos. However, the economic potential is often unclear. A sustainable solution requires a clear use case as in the example of cycle counting at IKEA.

1 Maghazei, O., Lewis, M., & Netland, T. (2022). Emerging technologies and the use case: A multi-year study. *Journal of Operations Management, 68*(6–7), 560–591.

13.4.3. Manufacturing operations in "finished goods" warehouses

In principle, finished goods are ready to ship, and all a manufacturer has to do is load them on a truck. However, in reality, what is usually referred to as "finished goods" in the plant still has hurdles to clear before getting out the door. The operations performed on finished goods at shipping typically include the following:

- **Boxing.** In most cases, packaging is done in production. In others, it is done as part of shipping. There are many different motivations for this:
  - Items are bulkier after packaging.
  - Packaging materials cannot be brought onto the production floor.
  - Customers have special packaging requirements.
  - Shrink wrapping: Pallets are often not shrink-wrapped until shipping, because otherwise it may have to be undone in order to ship half a pallet.
- **Recombining boxes.** Breaking down single-item pallets and recombining boxes into mix-loaded pallets.
- **Labeling.** Affixing custom labels on boxes and pallets.
- **Quality assurance.** Performing quality checks of varying scope as demanded by customers.

In the previous section, we called "semi-finished goods" items that are stored inside the plant in generic form, and retrieved for customization prior to shipping. But customization does not stop at the plant's shipping dock. More is done at multiple levels in the distribution network. Common examples include the following:

- Dealer-installed options in cars.
- Paint-color matching in a hardware store.
- Matching makeup foundation to a customer's skin in a department store.
- Cutting tubes or rolls of fabric to length.

The products that reach customers' hands have not entirely been made inside the factory. As a consequence, some level of production skill and product knowledge are needed in the distribution chain. Since, in the eyes of customers, the manufacturer is responsible for quality regardless of where customization takes place, the manufacturer needs to control how this work is done. Therefore, services requiring product knowledge are usually beyond the scope of 3PL or distributors. They are rendered either by employees of the manufacturer located in service centers next to customers, or by employees of *dealers,* who are trained and certified by the manufacturer.

Even then, however, the level of control cannot match the factory environment. It is simply more difficult to ensure that a spoiler is installed exactly the same way in 2,000 dealerships than at 1 assembly station. This constrains customization operations to be simpler and more mistake-proof than those carried out internally.

In some industries, products are *configured* by dealers rather than customized. For example, supplying an excavator with a particular bucket involves no modification of the machine, but requires having a range of buckets available, as shown in Figure 13.39. It also requires knowing which buckets are compatible with which excavators.

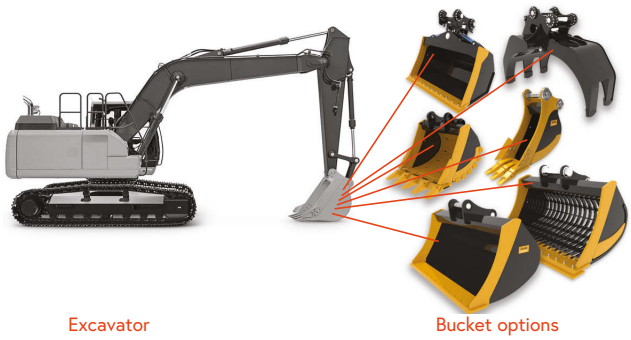


Figure 13.39 Excavator with bucket options

There are usually too many options for all dealers to have available at all times. The manufacture's Sales department then maintains a database of the inventory of buckets available at all dealerships, allowing dealers to locate where the nearest available bucket of a given type can be found, and the logistics of this activity include shipments between dealers.

13.4.4. Security and access control

Lead ingots may not need extensive warehouse security, but diamonds do. More generally, whenever items are light, easy to conceal, and saleable on the outside, warehouses need to be protected. The challenge is to provide the appropriate level of security without interfering with the flow of work, while treating employees with trust and respect. Most of the methods used to provide security generate more work or waiting time both for production operators and for materials handlers, as in the following examples:

- Entering the shop floor of some computer assembly plants requires going through metal detectors, in front of which there are long lines of employees at the beginning and the end of each shift.
- Some warehouse sections are surrounded by chicken-wire fences and accessed by opening a lock.
- Some items are distributed through vending machine-like dispensers that scan and validate users' badges, and track the quantities withdrawn.
- Like safety or mistake proofing, however, security should not add labor. The concern here is not just day-to-day productivity but vulnerability in a crunch. Any procedure that adds labor is likely to be bypassed under pressure. Only if a security procedure adds no labor can management be assured that it will be followed at all times, even when the work load is exceptionally high.

The following are a few approaches that management can use to increase security:

1. **Communication.** The need for security must be communicated to all employees through the management chain, from plant wide all-hands meetings to the daily start-of-shift meetings of every team.
2. **Visible management.** The lack of visible management makes it unlikely that disappearances will be noticed.
3. **Ease of execution of authorized actions.** Security procedures must be designed not only to prevent unauthorized actions but also to make sure that authorized, legitimate actions proceed unimpeded. This is also an area where auto-ID technology has a key role to play: On parts or bins, RFID tags can be used to detect unauthorized movements; on badges, to automatically open doors to authorized work areas for employees who previously had to key in a code, place a magnetic key on a reader, or use a key on a padlock.

4 **Employee discounts at company stores.** Where feasible, company products should be offered at deep discounts in employee stores. In a prestige cosmetics plant, for example, employees can buy obsolete promotional goods at prices that go a long way towards eliminating any temptation on regular goods.
5 **Tight inventory control.** If inventory is tracked closely, discrepancies will be flagged quicker, which lowers temptation of pilferage.
6 **Surveillance.** The last resort is CCTV cameras. Surveillance can have detrimental effects on morale and should only be used where managers see no other way.

13.5. Material stores inside the factory

In addition to warehouses for raw materials and finished goods, factories need intermediate stores for materials inside plants. There are three main ways to organize stores in plants:

- Central in-plant warehouses.
- Local supermarkets.
- Direct delivery from suppliers to the production line.

13.5.1. In-plant warehousing

We may need a central warehouse because of the way parts are received and consumed, but we should not expect this warehouse to perform better than small satellite warehouses spread throughout the plant.

In car assembly, some items go straight from the receiving dock to production and are never stored by the materials organization. In the system called "*body-on-sequence*," some model and option specific components are ordered when final assembly of the car starts. The supplier immediately makes and delivers them sequenced in truckloads that are conveyed to production without being put away in storage, as shown in Figure 13.40.

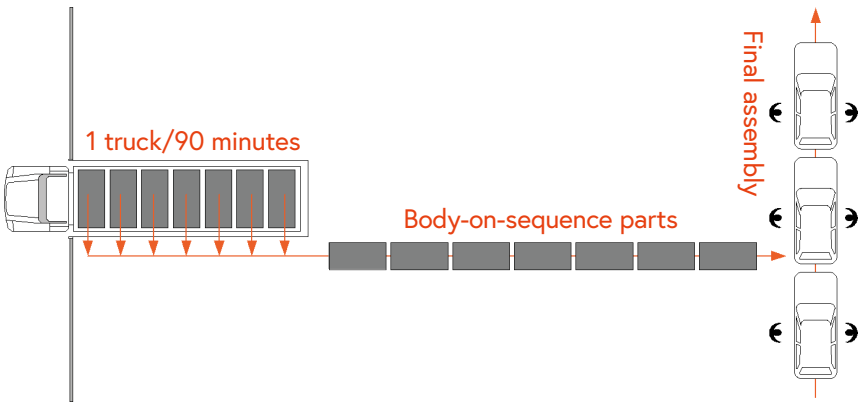


Figure 13.40 Body-on-sequence parts delivered straight to assembly

At the opposite end of the bill of materials are generic commodity items like nuts, bolts, and washers, which are provided on a *consignment* or *pay-per-build* basis by suppliers who not only manage the inventory but deliver the parts to the line side, so that the plant's materials organization never touches them.

Most items, however, particularly in industries other than car assembly, are neither managed by the body-on-sequence method nor by consignment. They arrive in quantities that the materials organization must receive, put away, and distribute to production.

The semiconductor equipment industry makes machines that sell for millions of dollars a piece and a high production rate for such machines is 1 unit per day. They contain custom-engineered mechanical parts as well as electronic controls that are delivered by currier services in box quantities. It is then feasible to split the content of these boxes at Receiving into plastic bins and immediately distribute these bins to small stores spread through the production area and referred to as "supermarkets," for picking by "water spider" operators. The disposable containers then do not make it past Receiving.

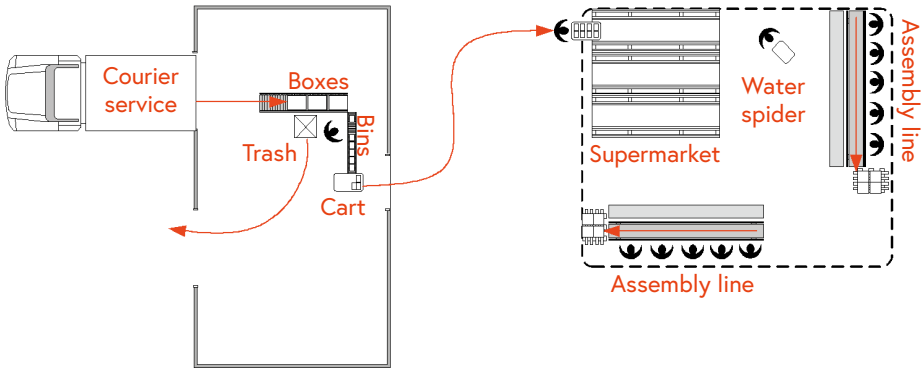


Figure 13.41 Delivery to a supermarket

With the higher volumes characteristic of the automobile industry, a similar situation can be achieved through the supplier milk runs with parts arriving in trucks that contain a mixed load from multiple suppliers with balanced quantities. It may then be possible to sort these items into quantities that can be directly delivered to *supermarkets* on the floor.

The supermarket is the interface between materials handlers and water spiders. The distinctions between these two roles is not always clearly understood, particularly among managers. A materials handler does the following:

- Move, store, and retrieve parts in boxes or bins.
- Rearrange boxes or bins.
- Switch parts between types of containers.
- Prepare containers for transportation by taping, shrink wrapping, fastening belts or rubber brands around pallets, etc.
- Print and attach labels, scan barcodes, and enter tracking data.

None of these tasks requires any product knowledge. Materials handlers do not need to know anything about the product other than outer dimensions and weight of its containers, full and empty. They don't need to know its bill of materials or manufacturing process. Water spiders, by contrast, don't have to drive forklifts, but they need product knowledge and they should be recruited among the most senior,

most skilled production operators. Their job is to prepare parts for immediate use in production, and includes the following:

- Removing any remaining packaging, plugs, or other protective hardware from parts.
- Feeding parts down narrow flow racks or other single piece presentation devices.
- Picking kits or matching sets of parts and delivering them to the line.

Clearly, the likelihood of having a kit with the wrong parts, too few parts, too many parts, or defective parts is substantially reduced when the picker knows the product and has experience assembling it. Before or instead of going to supermarkets, items with the following characteristics have to be stored in an incoming materials warehouse:

1. They arrive in minimum quantities that far exceed daily consumption, be it in the form of full truckloads consumed by the pallet or full pallets consumed by the bin.
2. They are used in multiple products whose demand volume and mix may change. Storing the items in a central warehouse postpones their pegging to a product – or family – and makes it easier to respond to changes in mix.
3. The supplier's deliveries are irregular, and a safety stock is needed to avoid shortages.
4. Due to quality issues, all the supplier's products are required to go through incoming inspection or testing.

Suppliers with subpar performance in any area are of course on notice that they must improve or they are being phased out, and you would expect that, over time, there would be none left. In reality, however, there will always be a few. New products with new technology requiring some debugging will have components that suppliers have difficulty making and that will need to be watched until they mature. Stellar performers will also lapse as a result of management changes, mergers and acquisitions, bad outsourcing decisions, and many other reasons.

13.5.2. Grouping items by source or by destination

Figure 13.42 shows different strategies for grouping items in materials stores.

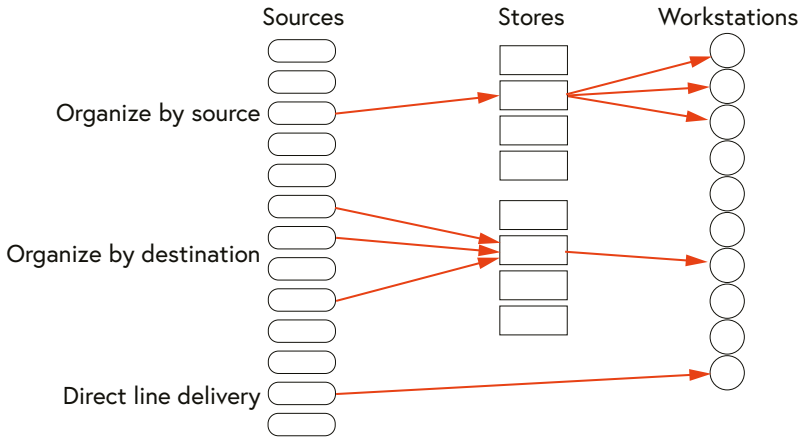


Figure 13.42 Item grouping strategies

As discussed above, *direct delivery* from Receiving to the line side as in the lowest path in Figure 13.42 is best when goods are delivered just-in-time or just-in-sequence, but that is rarely happening outside automobile assembly. When using intermediate stores, there are many ways to organize them, of which two are of particular interest.

In the upper path in Figure 13.42, the items are *grouped by source*, which is done in the central warehouse. The reasons for choosing this approach may vary. Car assembly plants typically have two materials service areas, one on each side of the assembly line. One is used to receive parts from local suppliers; the other from remote or overseas suppliers, whose lead times are substantially longer, and whose products may require customs clearance. If a supplier is on a watch list, it will be easier to monitor its performance if all its items are collocated in the warehouse. Shortages, overages, wrong items, and defective items will be easier to trace and respond to.

If only one item is on a watch list, as opposed to all items from a given supplier, it can be stored with other items going to the same destination, but must be flagged appropriately. We have seen one company hang diapers above racks containing such items, as a sign that they need "baby sitting" – that is, a detour through Quality Assurance before release to Production.

In the middle path of Figure 13.42 the items are *grouped by destination*, usually in a local supermarket near that destination. Parts from one source may go into different storage locations, but retrieval for one station is now one-stop shopping. Consolidation centers of retailers and logistics companies do exactly this job.

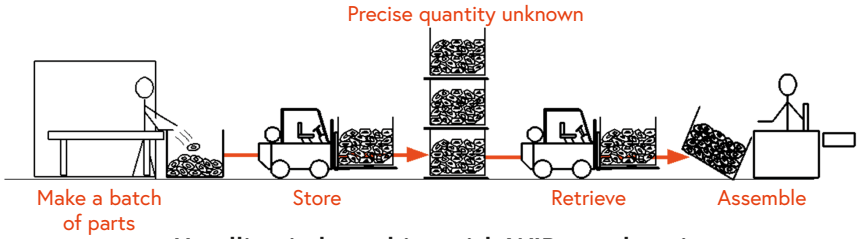
13.5.3. Stores for work-in-process and semi-finished goods

It is common in factories for work-in-process (WIP) to be stored in a warehouse between every two operations. After a batch of parts undergoes a turning operation, it checks into the warehouse, under the control of the materials group, and checks back out days or weeks later to go into drilling. This practice is unsound for several reasons:

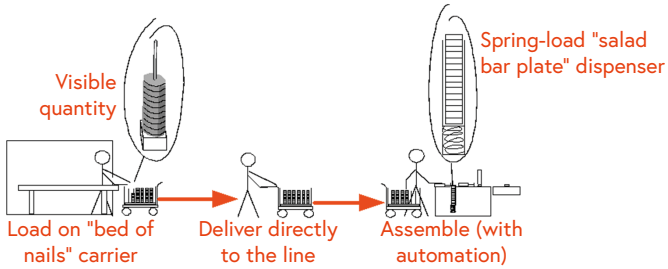
1. It generates unnecessary handling work, which not only wastes the time of materials handlers but also increases the risk of handling damage.
2. It reduces WIP visibility and, as a result, WIP quantities tend to grow.
3. It extends lead times.

There should be no warehousing of WIP by the Materials organization. It is held on the shop floor by Production, on lineside shelves, in input and output buffers around production lines, or in supermarkets for water spiders to pick from. The role of Materials should be limited to moving WIP between production lines, in response to pull signals. Figure 13.43 shows two ways of transferring parts from a fabrication machine or cell to assembly.

In Figure 13.43a, bagel-shaped parts are thrown into large bins, and put away in a warehouse prior to being delivered to manual assembly; in Figure 13.43b, they are neatly stacked on a custom cart that uses the parts' center hole to hold them on pins, and delivered directly to assembly. There, the operator transfers the parts, one column at a time, onto spring-loaded devices patterned after the plate dispensers used in salad bars, which presents the parts one at a time in a controlled orientation for automatic assembly.



a. Handling in large bins with WIP warehousing



b. Direct delivery in custom carrier with assembly automation

Figure 13.43 WIP warehousing versus direct delivery

There is a distinction, however, between WIP and semi-finished goods. It is common in many industries to make a small number of components or generic models, to be configured or customized just before shipping. The following are just a few examples:

- Cosmetics are compounded in bulk, and then the same compound goes into a variety of bottles, tubes, or jars that may further be aggregated into gift sets.
- Materials in the form of plastic tubes or sheets are first made in rolls, and then cut to custom dimensions as needed.
- Semiconductor wafers are fabricated into generic master slices that are burned into 100 times more different application-specific integrated circuits at the last operation.
- Commercial trucks are outfitted with customized interiors and painted with the customer's colors and logos.

Where such patterns are present, it often makes sense to make the generic item to stock and customize it to order. One consequence of this policy is that items that are neither incoming materials nor finished goods are warehoused and managed during that time by Materials rather than by Production. These items are often called "semi-finished goods."

The difference between WIP and semi-finished goods is in the eye of the beholder. What makes an item one or the other is not its physical nature but management decision. Returning to the example of a part that undergoes turning and drilling, the holes drilled into it may serve to mount it on the customer's product. If this part is mounted differently on multiple products, then drilling can be treated either as one step performed in a flexible machining cell, or as a customization step, with the part between turning and drilling handled as semi-finished goods rather than WIP. While storing semi-finished goods can reduce the cost of customization by supporting delayed differentiation, unless management is vigilant, this strategy can easily degenerate into a return to practice of warehousing WIP.

> **Definition**:
> Semi-finished goods are generic units made in much larger quantities than they are consumed in.

Table 13.1 shows possible patterns for a few types of semi-finished goods.

Table 13.1 Examples of storage and retrieval pattern for semi-finished goods

Type of semi-finished good	Arrival quantities	Picking quantities
Discrete pieces	Pallets, boxes	Bins
Film, plate, fabric, tubes	Rolls	Cut pieces
Liquid or paste	Drum, tank	Bucket

Semi-finished goods stores should exist only to support customization. The same item can be both WIP where used in its standard form in a high-volume product with a dedicated assembly line, and semi-finished when customized into multiple, low-volume products. In this case, the WIP fraction of its demand should be delivered directly to assembly, and only the rest stored in the semi-finished goods warehouse.

The semi-finished goods warehouse should also only hold items for which customization adds up to a steady demand for the generic item. If an item made in-house is consumed so rarely that it would not justify a dedicated location as an incoming material, then it should be excluded from the semi-finished goods warehouse. If a generic lot would last 6 months, the finished product should be made to order from scratch using a residual job-shop set aside for such random requests. These concepts are shown in an example in Figure 13.44.

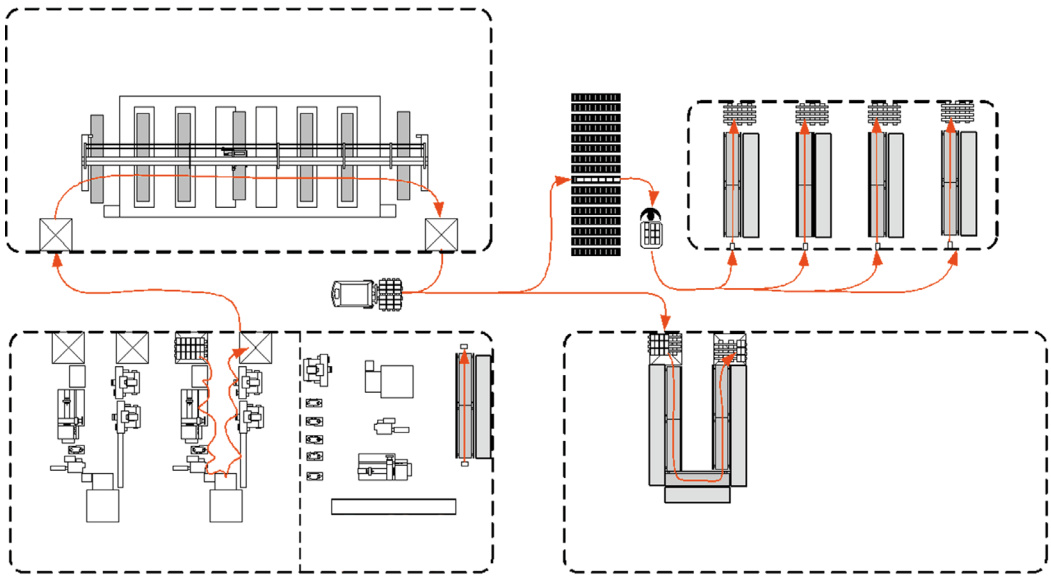


Figure 13.44 WIP and semi-finished goods

Summary

The most essential points in this chapter were as follows:

1. Warehousing is needed when there is a mismatch between supply and demand. It is particularly important in make-to-stock environments, but all manufacturers are likely to have warehouse operations.
2. Warehouse management principles and recommendations generally apply to a distribution center as well as a warehouse in a manufacturing plant.
3. Warehouse Management Systems (WMS) are useful software to help manage a warehouse.
4. Performance degrades catastrophically when a warehouse is too full. An ~85% maximum occupancy is recommended as rule-of-thumb, but the saturation point varies with the design and mode of operation of each warehouse.
5. Tracking materials into and out of a warehouse is essential to inventory accuracy. Parts in the warehouse should be touched only by materials handlers, withdrawn only with a pull signal from production, and the item and quantity withdrawn recorded, preferably automatically through auto-ID technology like barcodes or RFID. Parts should be in containers that make quantities visible.
6. Warehouse visibility is an essential best practice.
7. Parts should be stored according to frequency and volume of use, in addition to other logical product and supply chain characteristics:
   - The most frequently used items should have the most easily accessible, dedicated locations and use storage devices corresponding to their volume.
   - Infrequently used items are dynamically assigned slots like hotel guests.
   - Items from trusted suppliers exclusively bound for the same destination should be collocated.
   - Items from problem suppliers are easier to monitor if collocated by source.
   - Some parts and products require special storage and handling (e.g., temperature-sensitive products, explosion risk, heavy, bulky, or fragile products, regulatory or security issues).
8. There are many different racking systems. The three most usual ones in manufacturing are:
   - Standard wide-aisle, single-deep
   - Flow racks
   - Block stacking
9. Material handling and storage can be automated, but full automation of logistics is often not economical in manufacturing. Typical warehouse automation systems include:
   - Automated Storage and Retrieval Systems (AS/RS)
   - Carousels and storage lifts
   - Pick-by-voice and pick-to-light
   - Forked AGVs

- Special solutions such as AutoStore and Amazon Robotics.

10 Security is important but must not interfere with work and treat employees with respect.

Test yourself

1 Why do manufacturers have warehouses?
2 Should inventory be seen as a waste?
3 How full should a warehouse be?
4 How can visibility be increased in a warehouse?
5 What is inventory accuracy and how can you achieve it?
6 When is automation most applicable to manufacturing warehouses?
7 Can you describe five different warehouse automation technologies?
8 What is pick-by-voice and pick-to-light? What are the differences?
9 When should materials be transported from suppliers straight to the place of consumption in the factory and when should it pass through a warehouse?
10 How can you increase the levels of security in a warehouse?

Further reading

Frazelle, E. (2002). *World-Class warehousing and material handling*. McGraw-Hill.

Piasecki, D. J. (2003). *Inventory accuracy*. Ops Publishing.

Roser, C. (2020). Storage strategies. Series of blog posts on *AllAboutLean.com*

Richards, G. (2017). *Warehouse management: A complete guide to improving efficiency and minimizing costs in the modern warehouse*. Kogan Page Publishers.

Ten Hompel, M. & Schmidt, T. (2008). *Warehouse management*. Springer Berlin.

Van den Berg, J. P. & Zijm, W. H. M. (1999). Models for warehouse management: Classification and examples. *International Journal of Production Economics, 59*(1-3), 519–528.

14 Supply Chain Management

No factory is an island. Manufacturers are dependent on their supply chains – from the upstream raw material to the downstream consumer. These supply chains can be *managed*, at least to some extent, which has given rise to the discipline called "supply chain management" (SCM).

In reality, supply chains are not chains but complex networks that exist in larger ecosystems of industrial and social activity. While there is a rich debate on whether the right term is "supply chain," "demand chain," or "supply network," we chose a pragmatic approach and use the commonly established term SCM in this book.

A business idea from the 1990s is that competition is not between single companies but their supply chains.[1] While this insight can be helpful, there is only so much manufacturers can do to impact the supply chains they are a tiny part of. Since "managing" such complex networks is an impossible task for a manufacturer alone, focusing on the simple chain is more realistic. Most manufacturers have enough to deal with trying to coordinate with their direct suppliers and customers.

However, having an overview of the supply chain and a finger on its pulse can be a formidable competitive advantage for a firm. This chapter covers key tasks of SCM, seen from a manufacturer's perspective. Manufacturers must know how to design and analyze supply chains, how to contain the bullwhip effect, how to synchronize supply, how to develop suppliers and be developed by customers, and how to select outbound sales channels. Manufacturers also need to keep an eye on new developments such as the platform economy, and what it can mean for their operations.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Understand the tasks of SCM for a manufacturer.
2. Describe and analyze a supply chain network.
3. Implement strategies to tame the bullwhip effect.
4. Decide when vendor-managed inventory is a viable approach.
5. Involve in supplier selection and supplier development programs.
6. Decide what outbound supply chain channel to use for a product.

1 Christopher, M. (1998). *Logistics and Supply Change Management* (2nd ed.). Pitman.

DOI:10.4324/9781351110310-18

14.1. Supply chains and supply chain networks

The supply chain of a manufacturing company is the network of *other businesses* that provide it with materials, energy, and outsourced services on one side, and arrange for its products to reach end users on the other. Supply chain management (SCM) is the art of designing, negotiating, and running these external relationships that are vital as they are directly related to what a company spends and collects in revenue.

14.1.1. The scope of supply chain management

The term "supply chain management" was coined by consultant Keith Oliver at Booz Allen Hamilton in 1982. It was, however, first in the mid-1990s that the term became common place. Consultants and software developers then gave it broad definitions, similar to the one by Lambert and Cooper:[2]

> **Definition**:
> SCM is the integration of key business processes from end-user through original suppliers that provides products, services, and information that add value for customers and other stakeholders.

Working with other companies that operate independently, and over which management has no formal authority, is the special challenge that we call SCM. It includes coordinating with suppliers and customers.

In *internal logistics*, the entities that receive materials are commonly called "internal customers," based on the principle that "the next process is the customer." Internal suppliers, however, lack two key characteristics of customers: (1) They don't pay their internal suppliers, and (2) they don't have the option to switch suppliers. The statement uses a metaphor to say that you should treat the next process *as if it was a customer*, not that it actually is one.

In the supply chain, on the other hand, the links are true commercial relationships, that involve payments for deliveries and can be terminated by either side. It makes the supply chain a system in which outcomes are determined by the independent decisions of multiple entities.

At first, SCM was not covering much more than the buyer–supplier relationship, usually not stretching farther than from a company's direct suppliers and its direct customers. The supply chain was understood as a series of buyer–supplier links. Later, the term SCM has been expanded to include the whole "end-to-end" (E2E) chain from raw materials to consumption by end-users. In this view, a manufacturer's supply chain includes suppliers' suppliers, and their suppliers, and customers' customers – and their customers. Recently, the term "closed-loop supply chains," or the broader "circular economy," has been used to include the recycling or reuse of materials.

2 Lambert, D. M. & Cooper, M. C. (2000). Issues in Supply Chain Management. *Industrial Marketing Management, 29*(1), 65–83.

Box 14.1 **The SCOR framework**

Manufacturing professionals are likely to encounter the Supply Chain Operations Reference (SCOR) framework and should be familiar with the vocabulary it uses to define supply chains. SCOR is a process reference model developed by a consulting firm in 1996. It has been endorsed by the American Production and Inventory Control Society (APICS) and by the Association for Supply Chain Management (ASCM).

The SCOR framework in Figure 14.1 separates supply chain partners and five generic processes:

- *Plan*. Balancing demand and supply. The aim is to develop a feasible plan to meet sourcing, production, and delivery requirements.
- *Source*. Procurement of goods, parts, and services.
- *Make*. Manufacturing of goods, including internal logistics.
- *Deliver*. External logistics, including order management, transportation management, and distribution management.
- *Return*. Returning upstream or receiving downstream returned products. These processes extend into post-delivery customer support.
- *Enable*. "Managing" the supply chain, including agreeing on business rules, performance management, data management, resources, facilities, contracts, supply chain network management, managing regulatory compliance, and risk management.

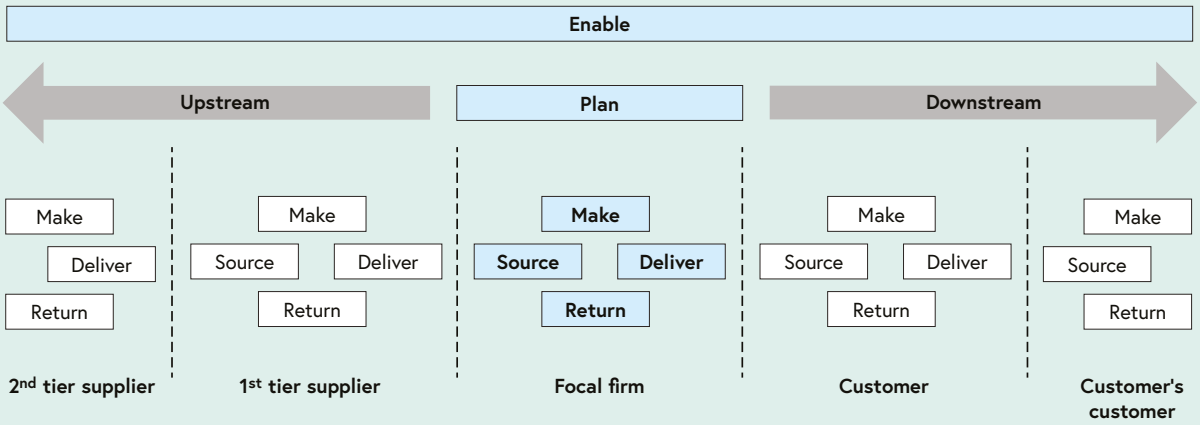


Figure 14.1 The SCOR framework

For each of these processes, the SCOR frameworks include high-level standards and "best practices" of conducting the processes. SCOR also offers suggestions for how to measure and follow up processes in the supply chain. As any standard, SCOR often struggles to cope with the complexity and variation at firm level. Yet, it offers a common vocabulary to discuss supply chain issues across the tiers in a supply chain. For the latest SCOR version, please check the webpage of the Association for Supply Chain Management: www.ascm.org.

14.1.2. **Vertical disintegration**

Ford's River Rouge plant near Detroit was set up in the 1920s as a vertically integrated plant that built cars from ore, sand, and other raw materials. By 1940, as shown in Figure 14.2, it employed about 100,000 people. In 2020, the Rouge is still operating but Ford only owns the assembly part of it. It is now an industrial park with a collocated supply chain, as opposed to an integrated plant.

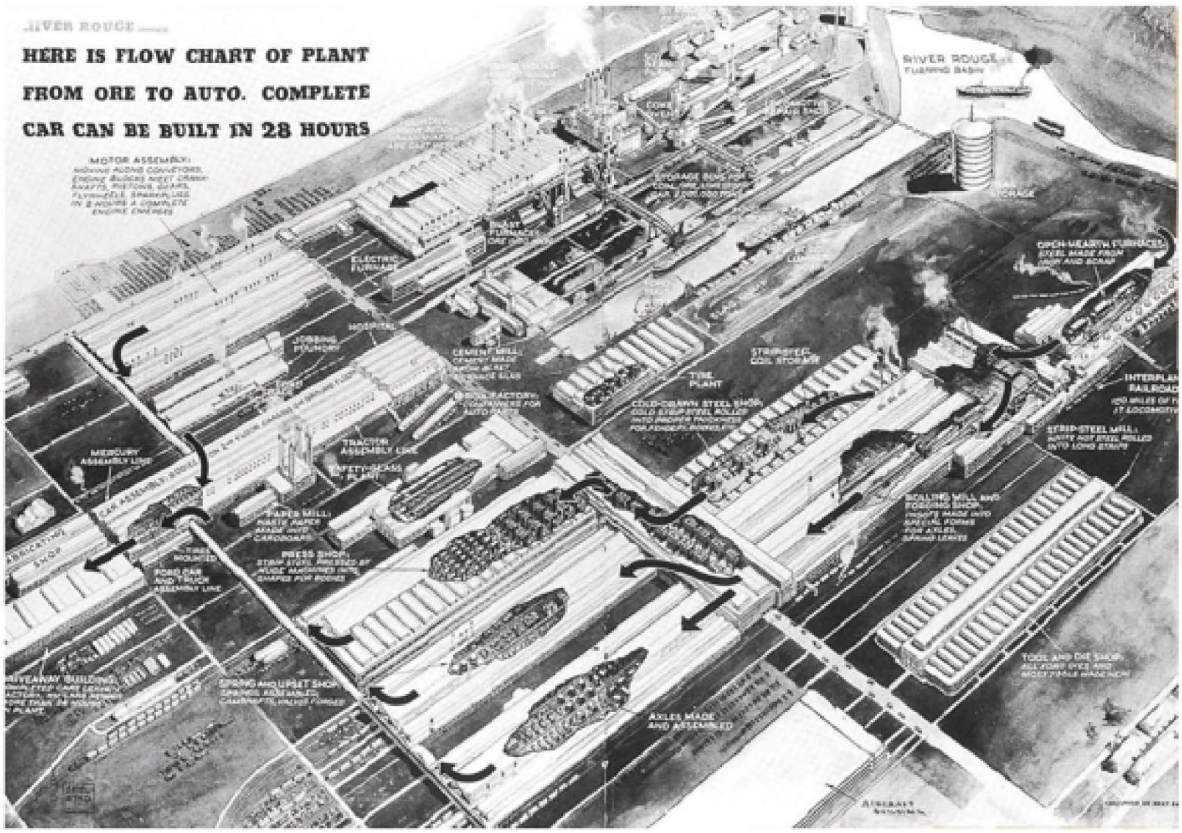


Figure 14.2 The Ford River Rouge Plant in 1940

When the semiconductor industry started in the 1960s, the chipmakers pulled their own silicon crystals and sawed them into wafers prior to making circuits on them, with production equipment that they built themselves. By the early 1980s, they bought silicon wafers and production machines from specialized suppliers. When an industry is new, vertical integration is the only option because there are no suppliers available. As it grows, the managers in charge of the end products tend to assume control of the company, leaving those in charge of components or equipment in a support position.

At a point in the maturation of the company, either the company spins off its component divisions or the managers and engineers of these component and equipment divisions may leave to start their own companies. Beyond this point, the internal divisions, if they still exist, struggle to stay competitive with external suppliers who have several advantages:

1. They spread the development costs of new technology over multiple customers.
2. They are run by managers who better understand and are focused on their specific business.
3. They are better able to attract new managerial and technical talent to this activity.
4. In many cases, like electronics, automotive, or aerospace parts, they have lower labor costs.

In addition, as an overall system, multi-tiered supply chains have proven that they are more flexible and resilient than a vertically integrated company. During booms, the customer company can grow without investing in capacity for components and machinery. During recessions, companies in the different tiers can adopt different survival strategies. To avoid layoffs, customer companies may bring some component production back in-house, while suppliers may diversify and enter markets that are less affected by the recession.

14.1.3. Supply chain coordination and orchestration

In reality, SCM is not really management. No firm *manages* the whole end-to-end supply chain. They manage their own activities within the supply network. The rest of management is really *coordination* activities between the firms.

Powerful firms like Airbus, IKEA, Procter & Gamble, Tesco, Lockheed Martin, or Walmart can take on a role that comes close to managing the supply chain. These firms dictate to their suppliers and sometimes their suppliers' suppliers the operational rules in the supply chain. The task of coordinating the whole supply chain, when left to one powerful partner, is called supply chain *orchestration*.

Why is SCM such a difficult task? There are problems both upstream and downstream.

- **Downstream.** A manufacturer's network of distributors and dealers often blocks the view of the market of end-users, which has driven the separation between a marketing function communicating with end-users and sales that interacts with direct buyers.
- **Upstream.** Suppliers, on the other side, are generally not keen to grant their customers access to their own suppliers. The supply network has a tier structure. The manufacturer buys from tier-1 suppliers, who buy from tier-2 suppliers, who buy from tier-3 suppliers, etc., all the way to a primary supplier who mines, harvests, or recycles basic materials.

Adding to these issues are complexities related to global production and logistics. Trade agreements, corporate tax policies, dynamic political environments, weather, war, pandemics, strikes, material shortage, and other local and global events add complexities to supply chain management.

Manufacturers can use various strategies to better keep their fingers on the pulse of the market, including selling directly to end-users, using a pull system with dealers, or setting up online support resources for end-users. To gain more visibility into the upstream supply chain, manufacturers can use purchasing power, if they have it, or negotiate reciprocal transparency and cooperation.

> **Stop and Think 14.1**
> During the COVID-19 pandemic, supply chains and SCM rose to the top of public awareness and was heavily debated in political discussions. Why? Why had it not been a "hot topic" before the pandemic?

14.1.4. **Mapping supply chains**

In order to understand the structure of supply chains, it is advisable to map and visualize them. Figure 14.3 shows a generic example of a supply chain.

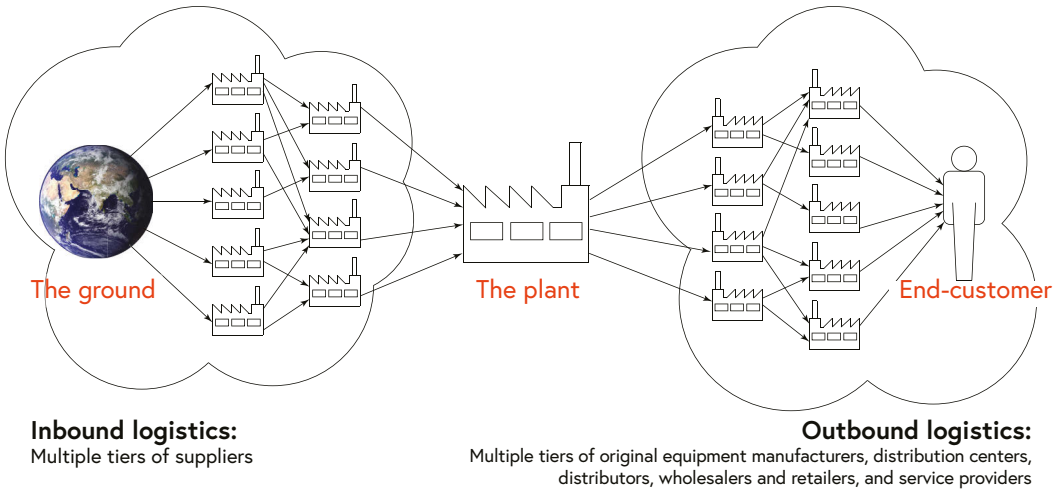


Figure 14.3 Mapping supply chains

In this regard, some common terms are useful:

- **Focal plant/firm**: The factory we are mapping for.
- **Vertical structure**: How "broad" the supply chain spans.
- **Horizontal structure**: The "length" of the supply chain from raw materials to consumption, which can be measured on number of tiers.
- **Horizontal position** (of focal plant/firm): The location of the focal plant in the horizontal network structure.
- **Downstream**: Moving to the right in the map means moving "downstream" towards the end-customer.
- **Upstream**: Moving to the left in the map means moving "upstream" towards the raw material.

Mapping a complete supply chain is a daunting task and often not necessary. Maps are summaries of a multidimensional reality drawn for a purpose, like navigating a road network or a subway system, and they omit the details. Figure 14.3 is intended to highlight the generic structure of flows of materials in a tiered supply chain, for the purpose of delivering goods. As such, it omits the backward flows of scrap recycling, product reuse, or waste disposal that are sometimes of economic importance. It also omits information flows.

It would be practically impossible to draw a complete map for a specific manufacturer with thousands of nodes, and unintelligible if anyone could draw it. Any mapping exercise calls for pragmatism. Maps are a support for spatial reasoning. They let the reader visualize events and flows occurring in parallel, and should be used where it is necessary to solve problems.

14.1.5. Materials and Information Flow Analysis

To solve problems in specific supplier relationships, Toyota came up with a graphic convention designed to represent flows of materials and data on the same map and called it "mono to joho no nagare" (物と情報の流れ), which translates to Materials and Information Flow Analysis (MIFA). If a link in the supply chain is malfunctioning, with late deliveries of wrong items in wrong quantities, then mapping the details of data exchanges and movements of materials provides a basis of discussion between supplier and customer, leading to a solution. Figure 14.4 shows an example of MIFA explaining the operation of the eKanban system in one link of the supply chain. This illustrates a micro link between partners.

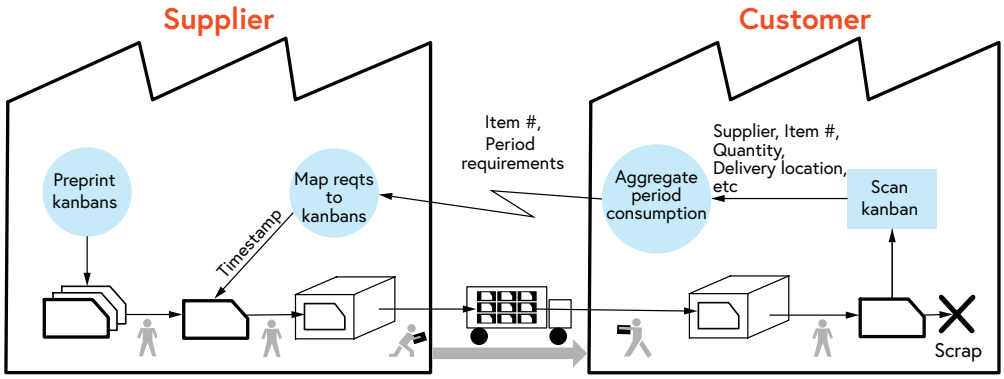


Figure 14.4 An eKanban MIFA

In 1998's *Learning to See,* Mike Rother and John Shook[3] rebranded the MIFA graphic convention as "Value Stream Mapping" (VSM) and promoted it as a tool to apply to all the production activity of a factory as a way to identify improvement opportunities. Rother and Shook were open about the position of this tool within Toyota, saying in the introduction that it was "never thought of as important in its own right," used "almost as an afterthought" and not as a means of learning to see.

In spite of these reservations, VSM took on a life of its own in Western factories, with managers mandating that these maps be drawn for every process and prominently posted. This has not happened in Japan, either at Toyota or at other company factories. To this date, the only Japanese publications about VSM are translations of American books. VSM is a helpful mapping tool when used strategically and not force-fitted. The simpler MIFA tool or a different notation that illustrates the case in point is often the better choice.

> **Stop and Think 14.2**
> Carefully review the simple MIFA in Figure 14.4, do you understand the main massage just by looking at the figure? What if the same message was explained in words?

3 Rother, M. & Shook, J. (2003). *Learning to see: Value-stream mapping to create value and eliminate Muda*. Lean Enterprise Institute.

14.2. Supply chain synchronization

Supply chain management is about getting the participants in your supply network to work towards a common end. With hundreds or even thousands of companies in different tiers, it is a daunting challenge. Let's examine some of the common dysfunctions in supply chains and their remedies.

14.2.1. The bullwhip effect

One might expect manufacturers of components or subassemblies that are upstream from the final assembly in the supply chain to have market-induced variations dampened by filtering through the final assembler's production planning and control system. In many cases, however, the opposite happens, and stable demand for finished goods translates into more and more erratic variations in orders to suppliers as one goes upstream in the supply chain (see Figure 14.5).

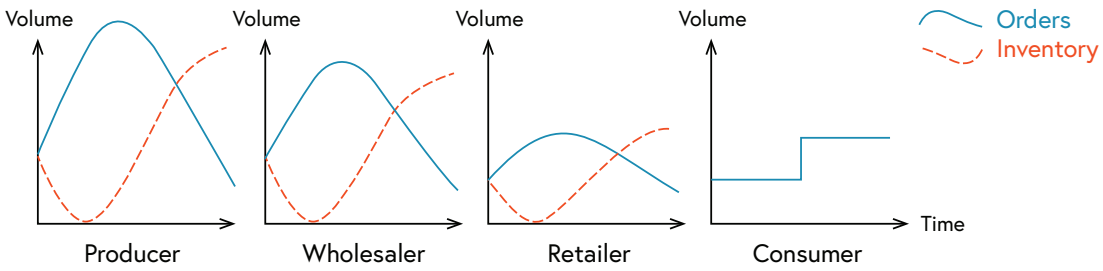


Figure 14.5 The bullwhip effect illustrated

Figure 14.6 exemplifies the size of the bullwhip effect in the textile industry a few decades ago, as described by George Stalk and Thomas M. Hout.[4]

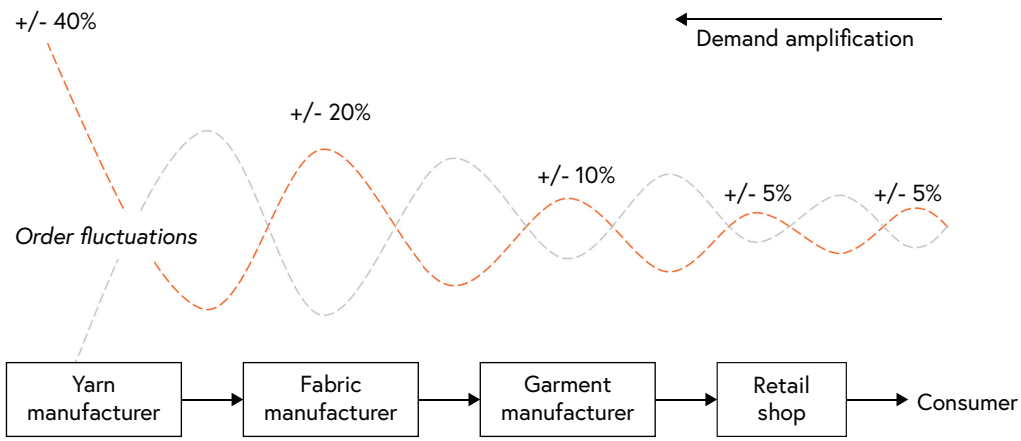


Figure 14.6 How the bullwhip effect artificially amplifies demand fluctuations in a supply chain

4 Stalk, G. & Hout, T. (1990). *Competing against time: How time-based strategies deliver superior performance*. The Free Press.

Hau Lee[5] called this the "bullwhip effect" and studied how the steady demand for diapers in the US market turns into alternations of feast and famine in the demand for the adhesive strips that hold the diapers in place. The same phenomenon can be observed in aerospace, where a demand for airliners that is steady and known months in advance turns into wild fluctuations in orders for components.

Earlier, Jay Forrester[6] had developed an example of how a seemingly logical method can generate the bullwhip effect. Let us assume that:

- For each link in the supply chain, we have a starting inventory at the beginning of the month, which is depleted by sales and replenished by purchases to a target level set at the end of every month at 3-months' worth of that month's sales.
- The demand is 100 units/month.
- We introduce a small perturbation in the sales volume of the last link and examine the response from the three preceding links. At the end of the 2nd period, the last link, the customer, is unable to ship the 100th unit in time to credit second-period sales for it but catches up the following morning. In Table 14.1, we see in the customer's "Sales" row for periods 2 and 3 that he sold respectively 99 and 101 units. At the end of Period 2, the customer, therefore, holds 300 - 99 = 201 units.
- Based on the logic described above, the target inventory is three times the periods' sales, which, for Period 2 is 3 × 99 = 297 units. To replenish the inventory to the target level, the customer, therefore, needs to order 297 - 201 = 96 units. At the end of Period 3, the same calculations applied to sales of 101 units lead the customer to order 107 units.

As we can see in Table 14.1, as we go up the chain in the following columns, this small perturbation in sales is amplified at each link by the planning logic, resulting in the plant ordering 36 units one period and 305 the next. When we get to the supplier, the fluctuations are so wild that it ends up not ordering at all in Period 5 and creating a shortage in Period 6.

Table 14.1 Emergence of the bullwhip effect

Supply Chain Link	Customer			Dist. center			Plant			Supplier		
Period	1	2	3	2	3	4	3	4	5	4	5	6
Starting inventory	300	300	297	300	300	288	300	300	252	300	300	264
Sales	100	99	101	100	96	107	100	84	140	100	36	264
Ending inventory	200	201	196	200	204	181	200	216	112	200	264	0
Target inventoy	300	297	303	300	288	321	300	252	420	300	108	792
Quantity ordered	100	*96*	*107*	100	*84*	*140*	100	*36*	*308*	100	0	792

While this example is contrived for illustration, the bullwhip effect is real and must be dealt with. Leveled scheduling in final assembly and the organization of milk run logistics with suppliers are effective countermeasures. They are, however, applied at the customer's initiative, and are no help to suppliers whose customers do not see a problem.

5 Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management Science, 43*(4), 546–558.

6 Forrester, J. W. (1961). *Industrial dynamics*. MIT Press.

A manufacturer can do the following:

1. Find out from the customer what the underlying stable demand for finished goods is.
2. Produce at this rate.
3. Maintain a finished goods inventory to absorb the fluctuations in orders.
4. Make the customer aware of the savings that can be achieved by eliminating these artificial fluctuations and the finishing goods stocks.

Whether the customer takes advantage of this opportunity or not, is of course not up to the supplier, but at least the supplier has made the consequences of the customer's policies visible.

Box 14.2 **The Beer Game**

The Beer Game is a serious game that teaches the dynamics of the bullwhip effect. It was developed in the early 1960s at MIT Sloan School of Management and has been used all over the world since then – and still is. The game is usually played as a group of four; playing the brewery, the distributor, the wholesaler, and the retailer. The game, originally developed as a board game, is also available in digital versions.

A game coordinator or a computer simulates the consumers' demand at the retailer. No communication is allowed in the supply chain during the game. The players only know their current stock, the inbound lead time (2 weeks), and the weekly customer order (including backorders). They also know the weekly cost of holding inventory and the cost of not being able to deliver customer orders (twice as expensive as holding inventory). Based on this information, they decide how much to order from their supplier each week.



Figure 14.7 Students playing a digital version of the beer game

The game is typically played for 40-50 weeks. Including introduction, the game can be played in roughly 1 hour 30 minutes. In addition comes setup time, which can be longer for the board game than a computer version. Importantly, there should also be time for reflection, explanation and discussion, which could require another 90 minutes.

14.2.2. Vendor-managed inventory

Vendor-managed inventory (VMI) is a method that helps supply chain synchronization of items that the customer does not feel a need to monitor closely. Commodities like nuts, bolts, and washers, are examples where customers like to delegate as much responsibility as possible to suppliers. As the name suggests, VMI is an approach in which the management of the stock of an item at a customer site is left entirely to the supplier, and it is not a zero-inventory system. It can be used in outbound logistics, where a manufacturer or wholesaler assumes responsibility for keeping the shelves of a retailer stocked, and in inbound logistics, where a supplier of components or materials keeps them in stock in the customer's factory.

In VMI, the supplier is given access to the customer's inventory database and is allowed to initiate shipments when reorder points are crossed. Consignment arrangements delegate more responsibility, by letting the supplier *own* the stocks in the customer's plant until the parts are used, and, in the *pay-per-build* or *consignment* scheme, suppliers are paid based on the quantities of their parts incorporated into finished goods leaving the production line, per the bill of materials.

The extent to which the vendor manages the inventory varies:

- The customer makes the data about the supplier's products in the inventory database accessible. The supplier monitors the quantities on hand and makes deliveries as needed to keep it within agreed-upon bounds. The materials from this supplier are received like other materials by the customer, and the supplier issues invoices together with advanced shipping notices (ASN).
- In addition to deciding when to replenish, the supplier takes charge of physically maintaining stores of the parts or materials on the customer's shop floor. Like for food vending machines in factories, supplier representatives are allowed onto the shop floor. The supplier retains ownership of the materials until they are used.

By definition, VMI schemes apply to items that the customer does not feel are worth managing. In a retail situation, the supplier's items may not be given great importance by a distributor carrying thousands from multiple suppliers. VMI then enables the vendor to give these items more attention and in particular to receive better feedback on which ones move or don't move. In inbound logistics, the approach is used in priority for commodities like nuts, bolts, and washers that may be used in large quantities but are inexpensive and easy to obtain.

Depending on the form it takes, VMI may or may not require extensive use of information technology. In the approach where the supplier remotely monitors quantities on hand, the supplier has to maintain an accurate inventory status database, making it accessible over the internet, and make it secure so grant authorized suppliers access to data about their items.

On the other hand, if the supplier simply keeps bins full of bolts on the customer shop floor and gets paid on a consignment/pay-per-build basis, then the *customer*'s information system does not need to monitor quantities on hand or issue orders in any form. All it needs is an accurate bill of materials to credit the supplier for parts in finished goods. We have seen cases where the adoption of consignment/pay-per-build for standard items reduced a customer's information systems costs by 10%, which can quickly amount to about $1M/year.

Box 14.3 **VMI at Barilla**

Italian pasta maker Barilla is known as one of the first examples of using VMI. David Simchi-Levi and coauthors[7] present the example of how Barilla in the 1990s was able, over a period of years, to change the ordering pattern of its distributors to what they called *just-in-time distribution* (JITD), so that Barilla would know in real time which products needed replenishment.

A key point of the story is that it took years for Barilla managers to convince both their own Sales department and buyers at one distributor to use the system. Once this approach caused sales through this distributor to rise, others joined in.



Figure 14.8 Pasta production at Barilla (Credit: Città di Parma/ Flickr, CC-BY-SA 2.0)

As of 2020, Barilla still used the same distributors and the approach was still called JITD. The team that manages it has grown and has enhanced it in the following areas:

- They have implemented new information systems integrated with production planning.
- They use point of sales information to improve sales forecasts.
- They have integrated JITD with business planning.
- The new technology has primarily been used to improve the existing JITD system, not to change the approach.

According to Claudio Ferrozzi, consultant to Barilla at the time, distributors were never happy to have others manage their stocks but accepted JITD because it was cost effective. In 2020, data is within everyone's reach, it can be processed in real time, and logistics companies can deliver within one day directly to retailers, working for dry products the same way as for fresh ones.

7 Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. (2008). *Designing and managing the supply chain: Concepts, strategies and case studies*. Tata McGraw-Hill Education. Chapter 4, pp. 67–82.

14.2.3. Dynamic replenishment

Pull systems are about replenishment, not order fulfillment. The signal to produce or deliver the amount of an item that has just been consumed is issued based on the assumption that the same amount will shortly be needed again, which is a naive, short-term forecast.

A naïve forecast simply assumes that the future will be identical to the present, and, under the right circumstances, cannot be substantially improved upon by more sophisticated methods. If the sun shines in a cloudless sky now, the naïve forecast that it will still do so in half an hour is about as accurate as a weather forecast can be. The accuracy of the naïve weather forecast, however, drops dramatically if the sky is not cloudless or the horizon lengthens to a day or a week.

In a pull system, the term of the implicit forecast is the replenishment lead time, which may range from 15 minutes to 1 week. This raises the following questions:

1. What are the constraints on the downstream process for the naïve forecast to be accurate?
2. Can more sophisticated forecasting methods improve the results?

In medium or long-range production planning, naïve forecasting is clearly insufficient, but pull system operations are not concerned with weeks and months but minutes and hours. In the automotive supply chain, dramatic changes in market demand do not occur within this time frame, but component suppliers and in-house subassembly lines on the receiving end of the bullwhip effect can be affected by changes in their customers' production schedules, which may be present even in the absence of any change in market demand. This is why pull systems go hand in hand with leveled sequencing.

A given number of pull signals can accommodate a range of activity levels simply by variations in their speed of rotation. In the automotive industry, the numbers of signals in a loop is reviewed periodically, every month or every other month, but is not changed by production control between reviews. During that time, production supervisors may withdraw signals from circulation to test the loop's ability to function with fewer signals, but this is not a response to a change in demand.

Other industries have different demand patterns. When a toy is a hit or an electronic device has pent-up demand, it is subject to explosive growth for a period of time. Figure 14.9 gives an example.

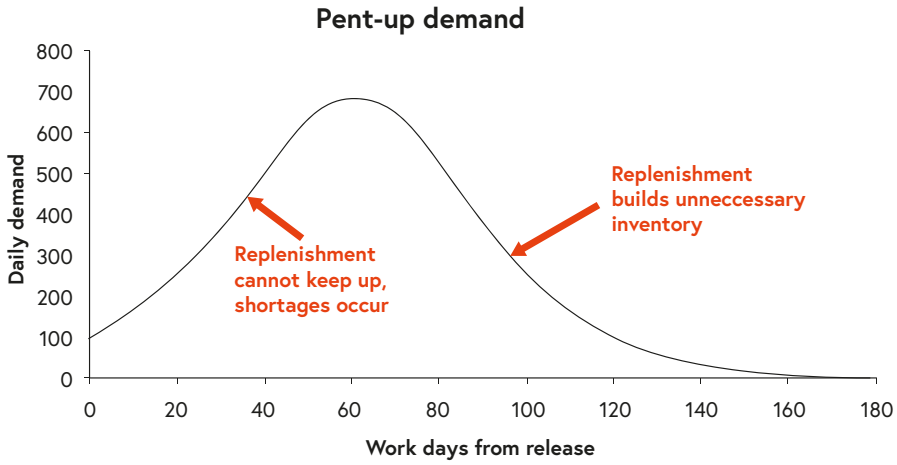


Figure 14.9 Pent-up demand pattern

> **Definition**:
> Pent-up demand refers to a situation where demand for a service or product is unusually strong.

Figure 14.9 shows pent-up demand where a population of buyers discovers that it needs a product as a result of its introduction. Until the product existed, they didn't know they needed it, but, as soon as it is introduced, they must have it. This pattern has been repeated over decades for products ranging from pocket calculators to smartphones. Demand explodes initially but then tails off as the market saturates. The curve in Figure 14.9 can only be drawn after the fact. As the product life unfolds, neither the height nor the timing of the peak is known.

This kind of demand is challenging in many respects, including the use of a pull system. A situation where demand doubles every month for three months is outside the domain of validity of Little's law. This type of variability is intrinsic to the market and not a side-effect of dysfunctional production scheduling.

In this context, a possible improvement over issuing signals to replenish exactly what has been consumed is to adjust dynamically either the number of signals issued or the quantities associated with each signal based on the trend in the demand. This means putting to use not only the immediate information on consumption but its recent history. Depending on how the pull system is implemented, this may or may not be possible, based on the following:

- If the signals are in the form of cards, bins, or other hardware devices, then their numbers cannot practically be adjusted every shift or every day, and neither can associated part quantities.
- If electronic signals are used, for example, with EDI, then dynamic adjustments can be made without creating more manual work.

Time-series analysis techniques to forecast demand in the near future based on the recent past are now available in inexpensive software for PCs, but possessing such software is not sufficient. A substantial effort is also needed to understand how to use it. Most lean manufacturing experts from the automotive industry view the development of what might be termed "dynamic pull" systems as a mistake, reflecting only a lack of understanding of pull systems. We view it instead as a legitimate effort to extend the concept of pull systems to domains where the demand does not have the same characteristics as that for cars and auto parts.

14.2.4. The body-on-sequence system

The body-on-sequence system is used by many carmakers under a variety of names, a few of which are listed in Table 14.2. This system is not common outside the automobile industry.

Table 14.2 Sequencing systems of different car makers

Carmaker	Sequencing system
Toyota	Just-in-time sequencing
Ford	In-line vehicle sequencing (ILVS)
GM	Just-in-sequence (JIS)
Chrysler	Sequence parts delivery (SPD)

The idea is to communicate the final assembly sequence to suppliers and require them to deliver parts in the same sequence, to the location and at the time they are needed. This is much tighter coupling of supplier and customer than is provided by VMI and the kanban system, and, at least in the version used by Toyota and GM, could not be implemented without computer networks.

Ford's ILVS system uses EDI to send sequences of vehicles to suppliers 6 days in advance. This means that either the system is used only for components whose sequence is not subject to change 6 days in advance or that last-minute alterations are made in the assembly plant, within the constraints of parts availability. Toyota also sequences parts from remote consolidation centers days in advance, but some model- and option-specific components made by local suppliers – like seats and upholstery components – are not ordered for each car until it starts at the first station of final assembly. Toyota, and GM, then use a system called "broadcast pulse" to send the requirements for one car at a time to the suppliers.

The reason this happens at the start of final assembly is specific to the car industry: The preceding operation is *painting*, which is the least reliable of the entire car assembly process. Just because the sequence from Production Control says that the next unit should be a white Corolla does not guarantee that there is a corresponding painted body available. On the other hand, once a car starts through final assembly, it can be counted on to make it all the way through, and to be, for example, at the seat installation station at a predictable time.

Suppliers who are new to this system initially respond by building finished goods stocks and only picking to the customer's sequence. Within a few years, however, they develop the capability to actually build in sequence, and the *supplier*'s assembly process is triggered by the broadcast pulse. At Injex in Hayward, CA, for example, the order fulfillment lead time was 220 minutes, from receipt of the broadcast pulse to delivery of the corresponding set of door panels to the proper dock at NUMMI, 14 miles down the road in Fremont. Injex held no finished goods stocks and its trucks made 10 round trips per day.

For this system to work, the supplier must not see large swings in its own part requirements within a shift. In other words, the body-on-sequence system cannot operate on the receiving end of the bullwhip effect, and a prerequisite is for the customer needs to sequence its products with the goal of smoothing the flow of components from suppliers.

The broadcast pulses from the body-on-sequence system *are* pull signals and are in fact the only type to eliminate any stocks between the supplier and the customer. In the supplier plant, there is work-in-process being assembled and a truck being loaded but nothing else. The logic of the body-on-sequence system is different from that of the kanban system, and it could not work without a computer network.

Stop and Think 14.3

If the body-on-sequence system helps automobile companies increase productivity, why isn't it implemented more broadly in other industries do you think?

14.3. Collaborative customer–supplier relationship

The following is a simplified rendition of a situation that commonly occurs in reality. A customer and a supplier are negotiating the renewal of a contract for a part. The customer is a large company that assembles consumer goods from components delivered by about 200 small to medium-size suppliers. Selling this part to this customer provides 20% of the supplier's revenue, who can therefore ill afford to lose it. The customer could switch to another supplier, but it would require a lengthy and costly qualification process and would involve additional risk. Failing to agree is therefore undesirable for both sides.

This kind of situation is common in the automobile or in the large appliance industry. The personal computer industry, by contrast, is dominated by suppliers of key components, like Intel, which are in a position of strength with respect to assemblers like Dell or HP. In the smartphone industry, Samsung is know as both a key supplier to Apple and other phone manufactures – and a competitor to them. Real negotiations in such settings involve many terms and conditions beyond price.

For illustration, we will assume that all non-cost issues have been settled already. The parties are therefore just negotiating the unit price to be paid for the part over a year, for volumes that are already agreed. The supplier's sales representative and the customer's purchasing agent both have limits they are not authorized to cross:

- The customer's upper limit is the price above which the customer will not buy. This may be the price paid to date for the part.
- The supplier's lower limit is the price below which the supplier will not sell.

While each party usually does not voluntarily reveal its limit to the other, we can assume that the following holds:

$$\text{Supplier Lower Limit} \leq \text{Customer Upper Limit}$$

Otherwise, the parties would not be talking. The goal of the negotiation is to agree on a price between these limits, and the payoffs for both sides are as follows:

- Customer payoff = Customer Upper Limit – Price
- Supplier payoff = Price – Supplier Lower Limit

This is summarized in Figure 14.10.

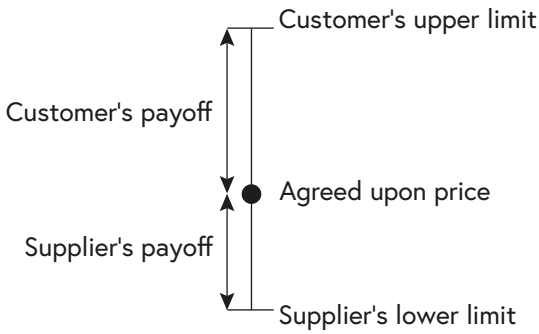


Figure 14.10 Negotiation parameters

14.3.1. The adversarial model

The following is, to date, the dominant model of customer–supplier relationships in the world. Each side is a black box to the other. As shown in Figure 14.11, the customer issues requirements specifications, orders, and quality problem reports; the supplier, lead times, advance shipping notices, and solutions to quality problems. The customer monitors the supplier's delivery and quality performance; the supplier, the customer's promptness in paying bills. They otherwise exchange no business or technical information. The customer maintains at least two sources for each component, both to ensure availability when one supplier fails, and to have leverage over each supplier in negotiations.

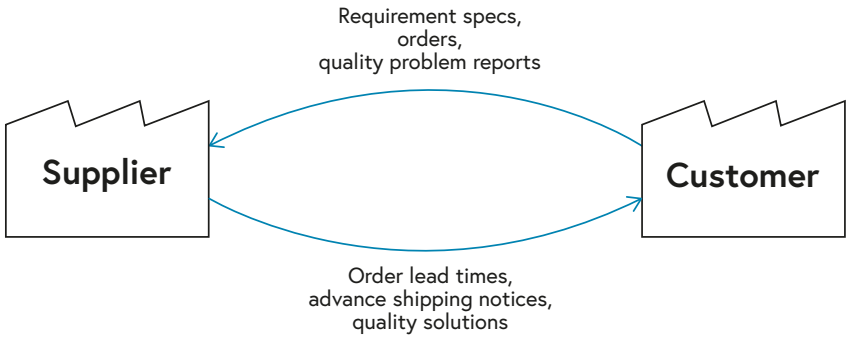


Figure 14.11 Information exchanged in adversarial mode

Neither side's limit can be moved, and the payoff for one side can only be increased through decreasing the other's by the same amount. Equally skilled and patient negotiators will eventually arrive at a price that is equally satisfying – or frustrating – for both sides. Since they do not know each other's limit, they do not know where the agreed-upon price falls between them, but they are in fact unlikely to split the difference at the midpoint between the upper and lower limit because each dollar payoff is not worth the same for both sides.

For the supplier, if the part is worth on the order of $10, an additional $1 in price represents another 10 percentage points of profit on it, and since the part is 20% of the supplier's business, another 2 percentage points of profit for the entire company. If the customer assembles the part that goes into a $1,000 product, then a $1 reduction in price only represents a 0.1% cost reduction for the product and is worth less for the customer than the supplier. Equal satisfaction, in this case, is achieved by a price that is closer to the supplier's limit than to the customer's.

14.3.2. The collaborative model

The collaborative approach to customer–supplier relations is in many ways the opposite. Instead of maintaining multiple sources and renewing contracts yearly, the customer commits to a single source for the life of the product and then works with the chosen supplier to make sure it is successful.

Instead of collecting bids for contract manufacturing of a part designed by the customer, the part is designed by the supplier's engineers, who may know of cheaper ways to meet the customer's requirements. The supplier's engineering and manufacturing capabilities are visible to the customer.

The customer and the supplier share business plans and forecasts, as well as technical information on production processes. The supplier may keep a representative in the customer plant observing assembly operations, and provide feedback on opportunities to make the part easier to assemble or less likely to be confused with other parts. Collaboration may also reveal that a tight tolerance the supplier cannot meet consistently is in fact unnecessary and can be relaxed. The customer provides "supplier support" including training and technical support on quality problem-solving and productivity improvement.

Consultant Kevin Hop[8] has summarized the differences between the two modes of operation as shown in Table 14.3.

Table 14.3 Differences between adversarial and collaborative supply chains

Organization	Adversarial approach	Collaborative approach
Top management, customer side	Not involved.	Meets with supplier top management on strategy.
Quality, customer side	Inspects at Receiving and issues problem reports.	Eliminates receiving inspection, rates suppliers, audits supplier processes, works with supplier on quality improvements.
Quality, supplier side	Replaces defectives and fills out problem reports. Little or no face-to-face interaction with the customer.	Works with customer quality on process improvements communicates issues freely, performs or pays for all inspections and repairs at the customer site.
Manufacturing engineering, customer side	Not involved.	Experts from customer plant migrate to supplier development group or are used on loan to support improving supplier processes in the area of quality, productivity and cost to the supplier. Improvements in cost. Savings are usually shared 50/50 with suppliers.
Product engineering, customer side	Design products and issues specs. Gives little support to design change requests.	Issues performance and functional specs to supplier, who designs the product. Open to design changes that benefit both parties.
Product engineering, supplier side	Weak or non-existent function. Maintains customer specs but has no input on improvements.	Strong function, with autonomous design capability, in partnership with the customer's product design group.

8 Personal communication

The result of this collaboration is that customers and suppliers have a bigger total payoff to share. Supplier quality improvements reduce the need for incoming inspection and testing and thereby make the parts more valuable to the customer. At the same time, improvements in the part's design and in the supplier's production process reduce the supplier's costs. Then the customer's upper limit goes up and the supplier's lower limit goes down, and *both* participants can have a larger payoff than with the adversarial approach, as shown in Figure 14.12.

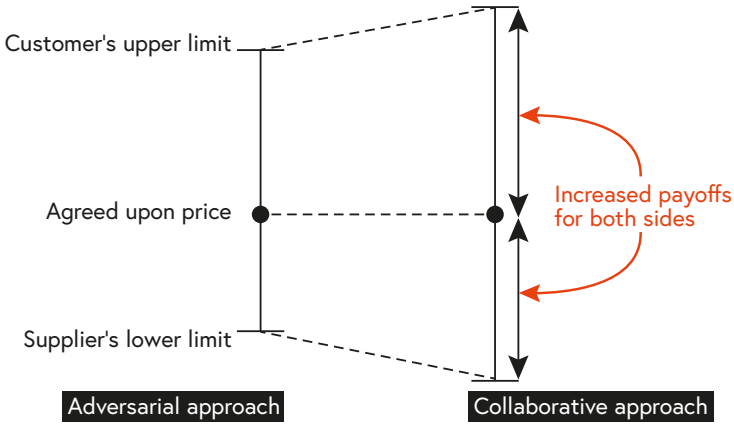


Figure 14.12 Adversarial versus collaborative approach

A collaborative supplier network has the following characteristics:

- **A small number of direct suppliers with a tier structure.** The customer relies on a tier structure allowing each large supplier to manage a group of smaller ones.
- **Single-sourcing.** The customers do not use the strategy of sourcing the same item from multiple suppliers to assure supply. The single-source suppliers may make their own second-sourcing agreements but they retain sole responsibility for the supply of the item.
- **Collaboration in product design.** The more complex a manufactured product is, the less sense it makes to treat its components like commodities. Instead, they are specific to the product and designed for it, and better, cheaper, and more manufacturable designs result when the engineering teams of supplier and customer collaborate during design in target costing, value engineering, and Design For Manufacturing and Assembly (DFMA).
- **Collaboration on cost reduction during production.** During production, supplier and customer work together to reduce costs through a process called *kaizen costing*, in which the customer provides technical assistance in exchange for price breaks.
- **Collaboration in problem-solving and emergency response.** Customers do not look for suppliers who "never have any problems" but for suppliers who don't hide them and are diligent about solving them, particularly when they are emergencies. If a key piece of equipment breaks down and cannot be repaired right away, the supplier notifies the customer immediately. If the breakdown was due to human error, a mistake-proofing device is implemented within a week.
- **A community.** The suppliers are an organized community. Suppliers to NUMMI participated in the Golden State Automotive Manufacturers Association (GAMA); suppliers to Toyota in Kentucky, in the Bluegrass Automotive Manufacturers Association (BAMA); suppliers to Applied Materials, in its Lean Suppliers Association (LSA). These groups hold conferences, visit each other's plants, and, in the case of Toyota, contribute to each other's improvement projects through activities called *jishuken* (自主研).

Many manufacturers are new to the idea of supporting suppliers. Since 1970, Toyota has undertaken to help its suppliers implement lean manufacturing. Since 1976, the main vehicle has been autonomous study groups (jishuken) that evolved in focus from information sharing to implementation of improvements and led to the development of *kaizen events* in the US.

Companies start supplier support groups once they have a track record of successful improvement in-house. They staff this group with experienced engineers who have been key contributors to the in-house effort. Successful groups make a sincere effort to understand each supplier's business, technology, and people, and do not just rely on compliance checklists to push suppliers into activities that have no bearing on their specific circumstances. The key objective is learning and improvement.[9]

Supplier support for quality assurance is often provided separately, but, in many cases, the most effective way for a supplier to improve quality is to implement lean manufacturing, and the two support efforts should be coordinated. Support should be provided to the entire business of the supplier, not only to production lines making the customer's products, and with a "bedside manner" that shows respect for the supplier's engineers and managers.

Suppliers benefit most from this support by taking the initiative of visiting customer plants, setting meeting agendas, and publishing minutes, taking advantage of training offered through the supplier association, presenting at supplier conferences, and maintaining communication with the customer's purchasing organization.

14.3.3. Transition to and stability of the collaborative mode

Given that it is beneficial to both sides, one might expect all industrial customers and suppliers to migrate promptly from the adversarial to the collaborative approach. It is not happening, and the collaborative approach that characterizes the Toyota supply chain remains the exception, at least in the US and in Europe.

Inertia is one explanation. The collaborative approach requires much more work to be put into a customer–supplier relationship, and today's buyers and sellers have not been trained to do it. They are heirs to generations that have practiced the adversarial approach and are comfortable with it. There is, however, another, more compelling reason: The switch requires both sides to change. The payoff matrix in Figure 14.13 shows what happens when only one side changes. They are playing a version of the prisoner's dilemma from game theory.

Both sides can choose to be adversarial or collaborative. Assume that both are initially adversarial, then for either side to switch to the collaborative mode when the other one doesn't benefits only the one that doesn't change. Any business or technical information it shares with the other side is used against it to obtain price concessions. In game theory terms, the situation where both sides are adversarial is a *Nash equilibrium*, because this strategy is optimal for each side, given what the other one is doing. The collaborative-collaborative corner is the best state for both parties, but it is unstable because both parties can make a short-term win by shifting to the adversarial state.

9 Powell D. J. & Coughlan, P. (2020). Rethinking lean supplier development as a learning system. *International Journal of Operations & Production Management, 40*(7/8), 921–943.

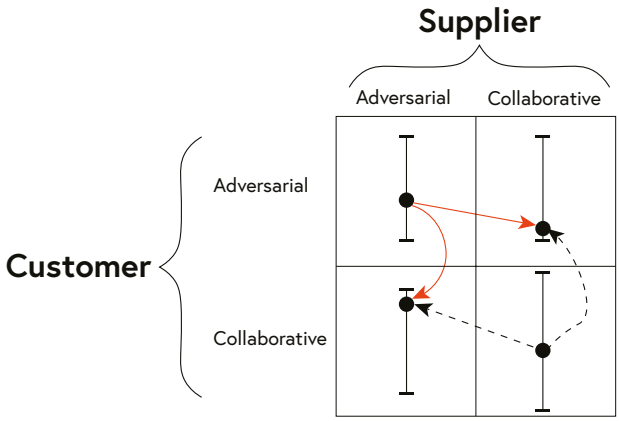


Figure 14.13 Payoff matrix

The only win-win move is for both sides to change strategies at the same time, and to grant each other a level of trust they have not had before. Since collaboration provides higher payoffs for both sides, one might expect that once achieved, it is stable. But it is not the case in Figure 14.13 above. Each side, assuming the other one remains collaborative, has the opportunity to gouge it by unilaterally turning adversarial. Although this move shrinks the total payoff, the side making it catches such a large share that it is advantageous, albeit for a short time: As soon as the other side catches on, we are back to the original mode where both sides are adversarial.

How does this play out in reality? A management change or an acquisition in the customer company may put in charge managers focused on cutting costs. First, they decide that supplier support is a luxury and gut it. Then they turn to suppliers, whom they feel have been coddled, and use all the information available to force them into price concessions. On the other side, a return to the adversarial mode initiated by a supplier will not take the form of an open attack. Instead, the supplier stops improving, takes advantage of its single-source position to stop investing in new technology, and uses the supplier support group's imperfect knowledge of the supplier's business to deceive it. Some observers in Japan blame Nissan's abuse by suppliers for the difficulties the company experienced in the 1990s, and credit the turnaround in part to the adoption of a tougher, more adversarial stance.[10]

14.3.4. **Co-opetition**

If few companies reap the benefits of collaborative customer–supplier relationships, it is not necessarily because their management does not see the value of it. Moving a customer–supplier relationship from the adversarial to the collaborative model is difficult because it requires a conscious effort on *both* sides. Furthermore, once it has been achieved, sustaining it over the long haul requires continuity in management on both sides and a determination not to pursue the short-term gains that can be obtained by a unilateral return to the adversarial model.

Working in a collaborative mode does not mean that price negotiations are a love fest. Collaboration enables the customer and the supplier to increase the total payoff of the relationship, but then they

10 Magee, D. (2003). *Turnaround: How Carlos Ghosn rescued Nissan*. Harper Business.

compete for their share of it in the price negotiation. In their 1977 book, Brandenburger and Nalebuff[11] call this behavior "co-opetition." In spirit, it is no different from collocating furniture stores to attract consumers to a "furniture district," and then competing for their patronage, or agreeing on telecommunication protocols and then competing in the market created for devices using these protocols.

The desire to work in this mode affects the way companies choose who they do business with. In selecting a supplier, for example, the customer may look more for openness and commitment to improving than for immediate performance. If, during the evaluation, a customer employee asks the supplier's manufacturing manager, "How do you solve your production problems?" "We don't have any production problems," is the wrong answer. It is defensive, not credible, and tells the customer that problems will be hidden. Every manufacturer has production problems, and what the evaluator wants to know is that the supplier will immediately notify the customer, and take short-term countermeasures while pursuing a lasting solution.

Professional supplier management will, in any case, require that the history of transactions with each supplier is summarized into supplier performance metrics. The metrics are used to certify some suppliers as trustworthy, others as needing help and guidance to achieve this certification, and yet others as candidates for replacement. The price suppliers charge is what purchasing traditionally pays the most attention to, among other reasons because price data is the most readily available. Collecting performance data on delivery and quality is necessary to keep the company from hurting itself by always choosing the lowest bidder.

14.3.5. Supplier support and development

Most manufacturing companies are used to providing customer support but new to the idea that they should in any way support or develop *suppliers*. In the collaborative mode of operation, customers and suppliers work together in the following areas:

- **Production system implementation.** Toyota supports its suppliers' efforts in this area, and these approaches have been adapted in other companies and industries.
- **Quality assurance.** While the production system improves quality, there usually is a separate organization in charge of supplier quality. In a collaborative relationship, the customer's Quality Assurance department does more than issue problem reports and demand additional record keeping and inspection. It also provides technical assistance in problem-solving.

Collaboration also extends to other areas that we will not go into here:

- **Product development.** In the system known as "black box" or "design in," customers issue functional requirements and engineers from the supplier and the customer organizations collaborate in the development of components that meet these requirements. This system is described by Takahiro Fujimoto[12] and by James Morgan and Jeff Liker[13] in the automobile parts industry.
- **New product introduction.** The new product introduction phase is the ramp-up from the production of sample quantities to the target production volume. Customers support suppliers in designing and running experiments to stabilize processes, and in setting up "run-at-rate" exercises in supplier plants, to enable the supplier to match the customer's planned takt time.

11 Brandenburger, A. & Nalebuff, B. (1996). *Co-opetition*. Currency/Doubleday.

12 Fujimoto, T. (1999). *The evolution of a manufacturing system at Toyota*. Oxford University Press.

13 Morgan, J. & Liker, J. K. (2020). *The Toyota Product Development System: Integrating people, process, and technology*. Taylor & Francis.

- **Information systems and production control.** Suppliers need training in understanding the business logic behind customers' information systems as well as daily transaction processing.

Box 14.4 **Toyota's supplier support system: Jishuken**

Toyota is not the only company in Japan or elsewhere to provide supplier support. Honda, Nissan, GM, Ford, and Chrysler all have organized efforts in this area. While not all aspects of Toyota's supplier support are transferable to other companies, industries, or countries, it is the reference. Toyota started a systematic effort to propagate knowledge of its production system among suppliers with the creation in 1976 of the "Society for Autonomous Research on the Toyota Production System." Its activities started with "Autonomous Study Groups," which is the literal meaning of jishuken (自主研). These groups met to discuss the nature of waste and ways to identify it. The term "jishuken" is still in use in the Toyota organization, with a much broader meaning, not only in Japan but in overseas transplant facilities as well.

The original idea was to bring together participants from non-competing companies to discuss the nature of waste and the means of recognizing it. This fit a pattern earlier established by the Japanese Union of Scientists and Engineers (JUSE) in propagating quality control concepts through study groups whose members could freely communicate because the other participants were non-competitors and low labor mobility in Japan ensured that they would not join a competitor. Inside companies, Quality Circles also started as study groups in the early 1960s.

Toyota's study groups rapidly expanded the scope of their activities, first to making proposals for waste elimination and then to actually implementing changes on the shop floor or in the support organization. By the mid-1980s, the focus of jishuken activity was to achieve immediate improvements while training staff to see waste, building teamwork, and improving communications. Today, Toyota sees jishuken as a structure to build the future.

Jishuken is intended to involve people in all levels of the hierarchy, as follows:

- *Team leaders and group leaders (first-line managers).* Workshops are held once a month, anyone can participate and is expected to. The team or group leader of the target area defines the problem but does not participate, and Maintenance is on standby to help implement changes.
- *Middle managers.* Workshops are also held once a month, but the participants rotate between plants.
- *Upper managers.* Higher-level managers rotate between supplier companies, working on larger-scale projects focused on innovation and involving multiple meetings over several months. This activity is part of business leadership development.

Of course, not everybody is always enthusiastic about interrupting their normal activities to participate in jishuken. To accommodate individual needs, the system provides flexibility both in the timing of their participation and in the format of the jishuken itself. The major risk in such a program is for it to degenerate over time into ceremonies where participants go through motions for form's sake.

Supplier support should not target *all* suppliers. It is more effective to have a deep impact on a few key suppliers than a superficial one on many. Toyota's OMCD works with about 50 tier-1 suppliers, out of the several hundred Toyota has in Japan. Then the support staff should invest its time in understanding the supplier's unique business rather than apply a one-size-fits-all approach.

Support should be provided to the entire business of the supplier, not only to production lines making the customer's products. This is needed for supplier personnel to trust that supplier support is actually intended to help them, as opposed to finding information to use against them in negotiations. Economically, only the customer receives the agreed price reductions, while cost savings on lines working for competitors add to the supplier's profits.

The suppliers receiving support may already have improvement programs in place. Supplier support should leverage, not ignore, these efforts, and show respect to the individuals involved.

14.4. The outbound supply chain

The bulk of the literature on supply chain management is devoted to the *inbound* side and managing relationships with suppliers. Strictly speaking, outbound supply chain management could be called *Customer Relationship Management* (CRM) but the software industry has monopolized this term for systems supporting sales, marketing, and service.

> **Definition**:
> The outbound supply chain is the multi-tiered network that products flow through from the manufacturers' finished goods stores to their end-users.

Like the inbound network, its management involves both physical distribution and information. The manufacturer needs not only a network of distribution centers and transportation services, but also the data about their activity. This data triggers daily actions, detects changes in the market, and takes them into account in production planning and scheduling.

14.4.1. Downstream distribution types

The structure of distribution matters to manufacturing. Physically, it shapes packaging, labeling, and transportation operations but, more importantly, it affects customer relationships and access to market information.

In a manufacturing company, Production has a single customer: Sales. Production's task is to make what Sales orders, regardless of whether it is based on orders, forecasts, or promotions. The *Sales & Operations Planning* (S&OP) process sets a framework to ensure that Sales does not place orders that cannot be delivered based on capacity, lead time, or quality.

Sales, in turn, takes orders from the 1st tier of the outbound supply chain, which may include end-users, dealers, distributors, or online marketplaces. These orders contain information about the demand volume in terms of trends, fluctuations, and breakdown by product, and reflect major events in politics, the

economy, or nature. The tiers in the outbound network alias filter and delay this information, possibly impairing the manufacturer's ability to respond.

Direct sales keep a manufacturer in contact with end-users, at the cost of holding all of the finished goods inventory, maintaining the sales and service infrastructure, and having the retail sales force on staff. Relying on multiple tiers of intermediaries reduces the direct costs of sales but increases the risks of misreading the market. Companies that go this route rely on their Marketing departments to compensate for lack of direct access to customers. In the following paragraphs, we review the choices made by several prominent companies in their distribution strategy. We separate five channels:

1. Direct sales
2. Dealers
3. Distribution/wholesale
4. Selling through platforms like Amazon or Alibaba
5. Omnichannels

Direct sales

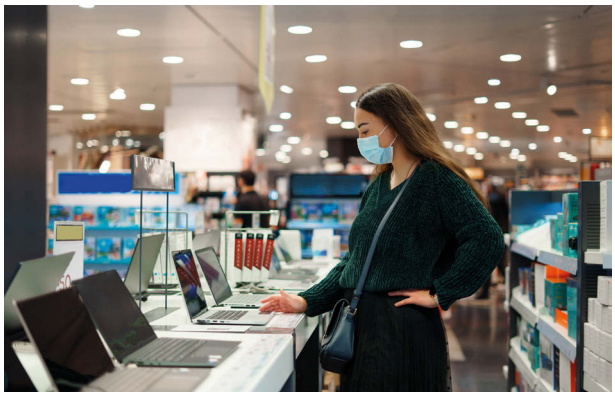
Direct sales have always been more common in business-to-business (B2B) than in business-to-consumer (B2C) activities. Airliners, production machinery, and high-end medical equipment are commonly sold directly by the manufacturer to customers but consumer goods usually are not. However, there are exceptions.

A straightforward example is "factory outlets." They are usual for confectionery products, household items, sports equipment, apparel, and similar. For manufactures of souvenirs or local handicraft, sales in local factory outlets can be the major share of their turnover. For most manufactures, however, direct sales through factory outlet channels are often only a tiny fraction of the total sales and have done more for brand building and sales of obsolescent inventory. Loyal customers come for a "good deal" or unique items. Figure 14.14 shows the example of Haribo's factory outlet at its headquarters factory in Bonn, Germany.


HARIBO
HARIBO
HARIBO

Dell has been selling computers directly to both consumers and business since it started in 1984, first by phone and then online. It used direct sales exclusively at first, but later allowed other e-commerce sites and physical retailers to sell its products.

Apple has been selling directly to end-customers since 2001, both through its website and its own chain of more than 500 Apple stores. Apple also sells through other retailers known as Apple Authorized Resellers, which includes Amazon among many others. At the time, except for Dell, computers were primarily sold in big-box electronic stores where products from multiple suppliers were side by side and the sales staff had limited knowledge of each model. Figure 14.15a and b shows some of the differences.



a. Laptop aisle in big box store



b. MacBooks in Apple store

Figure 14.15 Differences between Big Box and Apple Stores

Worldwide, Apple has located their brand stores in prestigious retail locations, like Fifth Avenue in New York City, the Ginza in Tokyo, or Gangnam in Seoul. When Apple started its stores, the big-box electronic stores had turned desktop and laptop computers into commodities. Steve Jobs wanted Apple products to stand apart and use a knowledgeable sales staff to promote their special features and justify higher prices. Selling directly to end-users also yields unfiltered and timely market information. The success of Apple's physical stores defied analysts expectations and led other companies to copy them, including Microsoft.

Toyota, in its early days, sold cars door-to-door to keep a finger on the pulse of the market. Chatting with housewives, sales reps from dealerships learned that a husband had just been promoted, and paid him a visit in the evening to sell him a new car to fit his new position. While it provided better market intelligence than dealerships, it has been abandoned, at least in urban areas. It was common practice in the 1960s and 1970s when not everyone had a car, and brand loyalty was not well established. It was less, "Would you like a Toyota today?" than following up with existing customers to keep them buying a new Toyota every 8 to 10 years, or upgrade, or buy one to accommodate a child, etc. It is still common in rural areas.

When buying a new car in Japan, traditionally the customer selected options with the sales representative and the company then built the car with those options. In the US, on the other hand, the dominant model is to buy from a dealer's inventory. Japanese customers buy used cars off the lot but rarely new ones. Japan has been slower in adopting online shopping than the US, but they have now, and with increased buying of cars online this has changed the behavior of sales reps.

Dealers

A dealer for cars, computers, or construction machinery is an independent business to which a manufacturer grants the right to sell and service products. Terms may vary but dealers are often the only outlet for the manufacturer's products within a given territory. They do not sell any competitors' products. The manufacturer trains dealers to sell and service products, and provides marketing and technical support. The dealers buy the products from the manufacturer and resell them at prices within guidelines from the manufacturer.

When Alfred P. Sloan[14] created the dealership system to sell GM cars in the 1920s, the primary purpose was to shield the production plants from the fluctuations in the market. Dealer inventories acted as a buffer to allow production plants to proceed at a constant pace. As he quickly discovered, however, this system also shielded GM from information about market trends, and they didn't realize the market had a downturn until the lots of their dealers were all full.

Today, as dealers report vehicle sales electronically, keyed on unique Vehicle Identification Numbers (VIN), car makers have access to sales activity in real time. Cars, however, are not the only industry using dealers. Furniture manufacturers like La-Z-Boy or Ethan Allen use dealers too, who may not provide this data as promptly.



Figure 14.16 Car dealership and furniture dealership

The dealer network contains finished goods inventory distributed across dealerships. When a customer asks for a product that is not in stock, the fastest way for a dealer to provide it is to get it from *other* dealers. Dealers, however, do not usually share their inventory data the way retail store chains do. If a chain store does not have in stock the espresso maker a customer wants, the sales representative looks it up in the chain's database and directs the customer to a store down the road that has it. Dealers, however, are independent businesses without a shared inventory database unless agreed and implemented.

Traditionally, sales reps would call other nearby dealers by phone and ask them for the product. With IT based on the *publish-and-subscribe* model, a dealer can publish the request to the dealer network, and other dealers' systems can automatically respond with bids to supply the product if they have it in stock.

14 Sloan, A. (1990). *My years with General Motors*. Currency.

Wholesalers or distributors

"Wholesaler" and "distributor" are often used interchangeably. A wholesaler buys products from a manufacturer but may work with competitors as well and does not provide service beyond resale. If your products are pasta or soap, you are more likely to use distributors than dealers. Wholesalers are a major players in the economy, but still unknown to the public. They often have unfamiliar names like *Avnet*, one of the world's largest wholesalers of electronic components, *Wuchan Zhongda Group Co., Ltd.*, distributing metals, ores, chemicals, rubbers, and more, or *Sysco Corporation*, the largest wholesaler of food and vegetables in the US.

Wholesalers/distributors buy goods in bulk from many manufacturers or other sellers and specialize in the sale of goods in bulk to anyone other than a standard consumer. For example, in pharmaceuticals, it is not uncommon that drugs are distributed by wholesalers to retail/pharmacy chains. Construction companies are served by wholesalers of tools, piping, and other building materials. In many industries, it is valuable for the customer to buy from a one-stop-shop wholesaler rather than managing many relations to different sellers.

Traditionally, wholesalers were located close to the markets they served. With e-commerce, wholesalers reach global markets even when located close to the manufacturing base, particularly China. For some products, like food and vegetables, wholesalers and their customers come together at specific wholesale market places.

Selling through platforms like Amazon or Alibaba

Online trading has been redefined by e-commerce firms such as Amazon and Alibaba. These giant firms are creating new opportunities for every firm to sell products directly to consumers via their market platforms. It has never been easier to reach millions of potential buyers, but not without the broker taking a share of the transaction. The platforms charge account fees and a percentage of each sale.

At **Amazon** the sale referral fee ranges from 6% to 45% of a product's selling price in 2022. The average seller pays about 15% per sale. In addition comes fulfillment and shipping cots, which can be self-managed or outsourced to Amazon for additional fees. While these numbers may sound high, sales and order fulfilment is an expensive activity left to the firm alone. It's a reason companies like Apple can be found on Amazon.

Some manufacturers, however, strategically decides against selling through Amazon. Examples are luxury labels Rolex and LVMH and furniture giant IKEA. German footwear manufacturer Birkenstock actively quit selling through Amazon when counterfeit products were also selling on the platform. All these brands can still be found on Amazon, but they are not offered by the OEM.

> **Stop and Think 14.4**
> Why do you think even famous manufactures with established selling channels offer their products on third-party platforms like Amazon or Alibaba?

Omnichannels

Omnichannels do not refer to a distinct channel but rather the coordination between many. Omnichannel retail is an advanced multichannel approach to sales. Multichannel retail allows customers to purchase products in many ways, without integration. Omnichannel retail seeks to provide the same experience across all commerce platforms. It focuses on providing a seamless customer experience whether the customer is shopping online from a mobile device, a laptop, or in a brick-and-mortar store.

For example, manufacturers can set up brand shops where products can be experienced, but where ordering can also be done fully online afterwards. There are many examples of firms successfully using an ominchannel sales strategy, including Amazon, Apple, Disney, IKEA, Nike. For example, both IKEA and Nike have developed apps consumers can use when browsing their physical shops, and immediately or later purchase items from. The latest development includes Augmented Reality (AR) technologies that let consumers "try on" shoes and furniture by using AR apps on their smartphones.

14.4.2. **Staying in touch with customers**

In addition to the standard marketing techniques like focus groups, surveys, and online reputation management services, manufacturers use a variety of approaches to sustain communication with the end users of their products around a multi-tier outbound supply chain. Some special examples include:

- **Web portals.** Many firms have customer loyalty programs, bonus cards and competitions that require a profile on their webpage. In addition, many modern products can be tracked directly in this portal, varying all the way from baby phones to cars. This way manufacturers can keep in touch with old and new customers and pick up changes in their lives and contexts.
- **Apps.** Apps can simply be small, mobile web portals but they can also be used to add more useful services to products. One example is lightening company Philips' Hue smart lightening system that allows customers to control their lightning from the phone. This allows Philips to access user data directly.
- **Social media.** Different social media platforms such as LinkedIn and Facebook have become standard platforms for customer contact and communication.
- **Product delivery at the factory.** For example, Porsche in Leipzig lets buyers pick up their Panameras and Cayennes at the assembly plant. The customers tour the spotless final assembly line, get an hour of coaching with a pro on the test track, eat at the fine-dining restaurant in the visitor center, and buy souvenirs. And, for this privilege, they pay an extra 1,250€.
- **Technicians at dealerships during product launch.** Product launch is a critical time, during which manufacturers need to respond promptly to the first customers. When selling through dealers, one method is to place technicians in residence in dealers' Service department, tasked with documenting and immediately reporting complaints or issues.
- **Customer site visits.** In a business-to-business (B2B) environment, and particularly when users of the products are in-house, it is often possible to organize site visits. For example, Boeing has a machine shop in Portland, OR, making structural components for aircraft. Once a quarter, each of its departments sends a small group of employees to the assembly plant that used its products to meet with the assemblers and collect their feedback. They record the interaction on video, and play it back to the entire production team back home. Different people go every quarter. It is relatively easy to organize such visits for an in-house supplier, to a plant that is not too far away. It is more complicated when you are dealing with actual customers, especially when the customers are not end-users but dealers or distributors.

14.4.3. The receiving end of supplier support

As earlier in this chapter, supplier support in B2B is usually described from the perspective of the customer company but it also must be considered from the perspective of the supplier who receives it.

If you make components to sell to businesses who assemble them into their products (OEMs), they may have policies restricting their purchasing to certified suppliers and send teams of auditors to assess whether you are qualified enough to do business with. Passing muster with these auditors is a necessary condition to work with this customer, and you must present your plant in a way that allows the auditors to check all the marks needed to give you the right score. In addition, customers may engage you in their suppliers' association and provide you with consulting assistance in improving your operations.

Working with your customers' supplier support organization – or supporting your own suppliers – is a different process, requiring a deeper level of involvement, and it is not a matter of public relations for either side, and should not be treated as one. The customer provides free consulting to help the supplier increase productivity and improve quality. In exchange, the supplier reduces prices by a fixed ratio every year, calculated so that the improvements are of economic benefit to both sides.

The customer pays less, while the supplier makes more profits. It is a win-win, but not an easy system to set up and operate. It involves top management, engineering on both sides, purchasing on the customer side, and customer service on the supplier side. It is not run by Public Relations and it requires a level of trust between supplier and customer that is not easy to achieve and sustain.

Suppliers usually have more than one customer. In many cases, the same production lines are used for multiple customers. In automotive parts, suppliers often have dedicated lines making customer-specific products, managed differently based on the policies of each customer. One customer, for example, may allow the supplier to change the layout of production lines as long as the processes remain the same. Another customer may want to approve any change in the line, including bringing two machines closer together without altering in any way the processes performed on these machines. In such cases, the supplier's continuous improvement activities are usually more successful with the lines serving the customers that give the supplier the most leeway.

Suppliers may also find themselves on the receiving end of supplier support from multiple customers, leading them in directions that are not always consistent. Even when it is from a single customer, the value of the advice varies with the capabilities of the individuals who provide it and the customer's own achievements in manufacturing processes.

Generally, a supplier should neither slavishly follow advice coming from a large, prestigious customer nor ignore input from smaller customers. Instead, it should practice continuous improvement on its own terms, accepting valuable input from customers and make it a true collaboration by occasionally contributing ideas that customers can use. As customers and suppliers both benefit from supplier support, both sides have an interest in making it work smoothly.

Summary

The most essentials points covered in this chapter were as follows:

1. Supply chain management is the coordination of activities across individual firms involved in the production and supply of products to end customers.
2. Manufacturers usually focus on coordinating the simple chain from 1st tier suppliers to factories to the 1st tier customer.
3. Supply chain mapping provides an understanding of the actors and relationships in the supply chain.
4. The bullwhip effect is the undesired amplification of orders and inventories upstream from the supply chain.
5. Vendor-managed inventory (VMI) means that the supplier replenishes the inventory at the customer site. Digital technologies allow the supplier to monitor the inventory.
6. Supporting suppliers to improve their operations can be a win-win but is hard to implement and sustain in reality because it requires playing with open cards on both sides on the business relationship.
7. There are several forms of channels for the outbound supply chain, including direct sales, dealers, wholesalers/distributors, and selling through platforms. Omnichannels means a coordinated approach across sales channels.
8. Platform companies such as Amazon and Alibaba are offering quick access to enormous global markets.

Test yourself

1. What does the management part in supply chain management mean?
2. Why have supply chains become longer and more global?
3. What is the SCOR framework?
4. What is supply chain orchestration?
5. Explain how to identify and tame the bullwhip effect.
6. For what type of parts is VMI generally recommended?
7. What is special about the body-on-sequence (or just-in-sequence) system?
8. When are dealerships used for the outbound supply chain?
9. What characterizes industries with powerful wholesalers?
10. What is special about Toyota's supplier development program (jishuken)?

Further reading

Ayers, J. B. (Ed.). (2004). *Supply chain project management*. St. Lucie Press.

Ayers, J. B. (Ed.). (2006). *Handbook of Supply Chain Management* (2nd ed.). Auerbach Publishing.

Choi, T. Y., Li, J., Rogers, D., Schoenherr, T., & Wagner, S. (Eds.). (2022). *The Oxford Handbook of Supply Chain Management*. Oxford University Press, New York.

Cooper, R. & Slagmulder R. (1999). *Supply chain development for the lean enterprise: Interorganizational cost management* (Strategies in confrontational cost management series). CRC Press.

Fisher, M. L. (1997). What is the right supply chain for your product. *HBR*, March-April, 105–116.

MacDuffie, J. P. & Helper, S. (2005). Collaboration in supply chains with and without trust. In: Heckscher, C. & Adler, P. S. (Eds.). *The firm as a collaborative community: Reconstructing trust in the knowledge economy*. Oxford University Press. Chapter 10, 417–466.

Pagh, J. D. & Cooper, M. C. (1998). Supply chain postponement and speculation strategies: How to choose the right strategy. *Journal of Business Logistics, 19*(2), 13–33.

Simchi-Levi, D., Kaminski, P., & Simchi-Levi, E. (2008). *Designing and managing the supply chain: Concepts, strategies and case studies*. McGraw-Hill.

Sinha, A., Bernardes, E., Calderon, R., & Wuest, T. (2021). *Digital supply networks*. McGraw Hill.

Part 5

Enhancing Performance

DOI:10.4324/9781351110310-19



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

15 Managing Performance

Managing performance is a part of running any business. Many measurements are required by law but measurements also communicate goals and align the organization in achieving them.

Measurements help managers gain insight about current performance and, hence, facilitate corrective actions, improvement, and organizational change. For employees, having performance measures helps clarify success criteria in their jobs. It can increase motivation to perform well and increase self-esteem when the performance is good. It also provides them a benchmark for improvement. In short, the use of measurement is an integral part of managing companies. However, it is difficult to do right.

Peter Drucker is often quoted as saying, "You can't manage what you don't measure." Many managers agreed and now believe that employees will do whatever it takes to maximize their scores. Given flawed metrics, it is fortunate for companies that these managers are wrong. If they had been right, all the companies that measure productivity in terms of sales-per-employee would have outsourced all production. They didn't, because metrics are only one of many factors influencing behavior. Most employees, at all levels, will not maximize their metrics through actions they feel violate common sense or are inconsistent with their personal ethics.

This chapter introduces performance management, which involves the design of performance measurement systems, the choice of metrics, the policies for sharing, summarizing, and visualizing the data, and the management actions driven by the data. It is about daily management, its technical tools and management practices. It is also about managing performance in times of emergencies.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Manage with performance metrics without being managed by the metrics.
2. Develop a Balanced Scorecard with leading and lagging metrics.
3. Separate the language of things from the language of money.
4. Choose appropriate key performance indicators.
5. Create a shop floor performance board.
6. Implement and manage shop floor daily meetings.
7. Reduce opportunities for metric gaming.
8. Implement risk-reducing supply strategies.

DOI:10.4324/9781351110310-20

15.1. Introduction to measurement

In sports, your score or your rank in competition is, by definition, the measure of success, but managing a day-to-day manufacturing business requires more than a single number.

15.1.1. Measurements, metrics, and key performance indicators

In a factory, you *measure* parameters and you *observe* attributes. Measurements are numbers and attributes are characteristics that are either present or absent. Most measurements and attributes are produced by instruments from sensor input or by IT systems from transactions. Measurements, for example, include time stamps, pressure, temperature, or energy consumption; attributes, whether or not a part is defective or a machine down.

Performance management usually involves synthesizing the raw observations of measurements and attributes into statistics called *metrics*. For example, the quantity produced on a given day can be subtracted from the planned quantity as a measure of plan adherence. Frequently observed attributes are often summarized into numbers as occurrence counts and ratios. Rarely observed attributes are kept as series of events. Metrics are chosen to act as *indicators* of performance. A company can use hundreds of metrics.

Well-defined metrics are of quantities that exist in a model of the activity regardless of our ability to estimate them, like the speed of light in a vacuum or the length of a rod coming out of a cutting machine. We design a rod to have a specific length. The rod length measurement ia a first-level abstraction, just one step from the physical object, as in Figure 15.1.



Figure 15.1 Length measurements as approximations

When we pile abstractions on top of other abstractions to come up, for example, with unit costs for manufactured goods or with ratios like Return On Assets (ROA). The very existence of a mapping from these numbers to reality is not a foregone conclusion, and it's a valid question about every metric

that is *defined* by a formula. The answer is necessary to understand how a change in the metric affects the process and the business as a system.

Managers select a few metrics to use as *key performance indicators (*KPIs). "KPI" is one of the most commonly used abbreviation in business. The KPIs are focused on the few performance areas that are critical to the organization's success. The word "key" means that you are not supposed to have many, and a number between 1 and 10 is reasonable (see Figure 15.2). Fewer is better, as long as the KPIs capture the essential information. What is "key," however, varies with the process, the area, and the management level. Inventory turnover, for example, can be a KPI for the Warehouse but not the Factory.

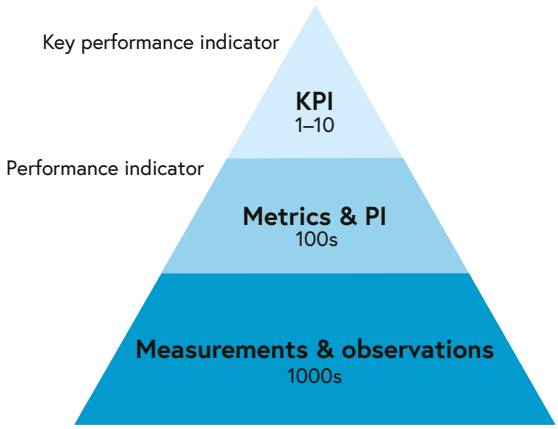


Figure 15.2 Hierarchy of indicators, performance indicators (metrics), and key performance indicators

Sometimes direct measurements can be KPIs. The temperature in cold storage can be a KPI for the warehouse and the number of people absent on a given day can be a KPI for HR. KPIs, however, are more commonly *metrics calculated* from several measured parameters or attributes. Not all metrics are "key."

15.1.2. Leading and lagging metrics

Leading indicators predict the future; lagging indicators report the past. A factory needs both. Different stakeholders want different metrics. A shareholder may be satisfied to know how well the company did last month and use this number as a predictor of future performance. A factory manager, on the other hand, needs to know how well the factory performs *today* and will in the next hours, days, weeks, and months. The same factory manager is also interested in how the factory performed in the immediate past, because it has career implications.

What ultimately matters is outcomes but we cannot manage a process based only on its outcome. The best way to achieve good outcome measures is to have good leading metrics. Consider the example of driving a car; leading indicators help us navigate the road ahead, while lagging indicators look backwards, through the rear window, at the road already travelled. The speedometer is a leading indicator of when we will arrive. The trip odometer is a lagging indicator, telling us how far we have travelled. The line between leading and lagging is, however, not always clear. The fuel gauge, for example, is telling us how much gasoline we have used – a lagging indicator – but it also helps us estimate our remaining range – a leading indicator.

Because it is much easier to measure results of a process, manufacturers are often stuck with only outcome measures. They reflect the work done while the plant was making the parts, not the work happening right now on the current WIP. As illustrated in Figure 15.3, when one drives a car by only looking in the rearview mirror, one will not get far.



Figure 15.3 Difference between leading and lagging indicators

Generally speaking, the closer you move to measuring process inputs and activities, the closer you get to leading indicators – of lagging performance outcomes. If you are measuring aggregated results at an organizational level, you are more likely using lagging indicators. The key to picking good leading measures is to understand the process that leads to the outcome.

15.1.3. SMART requirements for good metrics

There are requirements for good metrics. Back in 1981, consultant George Doran summarized a simple rule for making good goals: Good goals are specific, measurable, attainable, relevant, and timely (SMART).[1]

Below, the SMART acronym is applied to manufacturing metrics:

- Specific. A good metric measures what it intends to measure and is immediately understandable. No training or even explanation is required, and the number directly maps to reality, free of any manipulation. One type of common manipulation is to assume that one particular ratio cannot possibly be over 85%, and redefine 85% for this ratio as "100% performance." While this makes performance look better, it also makes the number misleading. Some companies calculate scores based on points awarded or deducted for a checklist of observations. This kind of highly processed data is *not* immediately understandable. Managers then often use scores to *rank* people,

1 Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review, 70*(11), 35-36.

departments, or companies. A rank only measures performance relative to a peer group, not in absolute terms. It is possible to perform poorly and yet rank #1 if all others do worse, which breeds complacency.

- **Measurable.** The input data of the metric should be easy to collect. Lead time statistics, for example, require entry and exit timestamps by unit of production. The difference between these times then only gives you the lead time in calendar time, not in work time. To get lead times in work time, you then have to match the timestamps against the plant's work calendar. Lead time information, however, can be inferred from WIP and WIP age data, which can be collected by direct observation of WIP on the shop floor. Metrics of WIP, therefore, contain some of the same information but are easier to collect.
- **Attainable.** People see how they can affect the outcome. With a good metric, each employee understands what kind of actions can affect the value. A shop floor metric, for example, should not be a function of the price of oil in the world market, because there is nothing operators can do to affect it. On the other hand, they can affect the number of labor hours required per unit, or the rework rate.
- **Relevant.** A better value for the metric always means better business performance for the company. This is perhaps the most difficult characteristic to guarantee. Equipment efficiency measures are notorious for failing at this, because maximizing them often leads to overproduction and WIP accumulation. Metrics should also have the appropriate sensitivity. If daily fluctuations are not what is of interest, then they need to be filtered out. A common method for doing this is to plot 5-day moving averages instead of individual values – that is, the point plotted today is the average of the values observed in the last five days. Daily fluctuations are smoothed away, but weekly trends clearly show. There are many other methods, depending on the nature of the metric, whether its current value is correlated with past values, and the accuracy and precision of the measurement methods.
- **Timely.** Metrics should be timely. There is a limited point in knowing the percentage of faulty units produced *after* they have been shipped to customers. The time dimension also emphasizes the importance of having leading indicators supporting lagging indicators. Leading indicators help users manage performance actively.

Stop and Think 15.1
Do you have a personal objective? Can the achievement of it be measured and followed up? Is the metric SMART?

15.1.4. The language of things or the language of money?

Shop floor metrics should be in the language of things rather than the language of money. The shop floor is a world where you speak in numbers of units made, time spent, or weight consumed, and it is several organizational layers removed from the boundary of the company where money changes hands. Metrics posted on the shop floor must therefore be nonfinancial. This does not mean that financials should be hidden from shop floor personnel, just that they should not be the basis for metrics these people review everyday.

Accountants translate the shop floor metrics in the language of things into the language of money for communication up the management chain. This is not to hide financial information from shop floor people but management should make it available on request rather than broadcast or post it on bulletin boards. Metrics based on financials are useful, but at a higher level of aggregation than the shop floor

and for other audiences than teams of operators. In some areas, it is also difficult to arrive at financial metrics that accurately represent the business impact of activities and events occurring on the shop floor.

Even at the overall factory level, companies should not look only at financial metrics, but consider multiple dimensions of performance instead. In *Balanced Scorecard*,[2] Kaplan and Norton proposed the set outlined in Box 15.1. The widely used DuPont model in Box 15.2 is from 1915 and was developed for top managers to compare the performance of divisions in *different businesses*, with only the language of money in common. Its numbers could be calculated with the information technology of the 1910s.

Box 15.1 **The Balanced Scorecard**

In the 1990s, Robert Kaplan and David Norton noticed that companies had a very unbalanced perspectives to measures. All companies focused on financial measurements, but far less on other capabilities. As illustrated in Figure 15.4, the Balanced Scorecard uses four perspectives to encourage a more balanced performance management:

1. *The finance perspective*. Typical measures here are earnings, revenue, return on investment, manufacturing cost, cost of poor quality.
2. *The customer perspective*. Typical measures are customer satisfaction, complaint rates, and market share.
3. *The internal business perspective*. This is where we measure the direct performance of our operations. We can track process quality, product quality, scrap rates, speed and efficiency, and a range of other metrics.
4. *The innovation and learning perspective*. These measures relate indirectly to today's operations but will impact how well we do it in the future. Examples include metrics for continuous improvement and skill development.

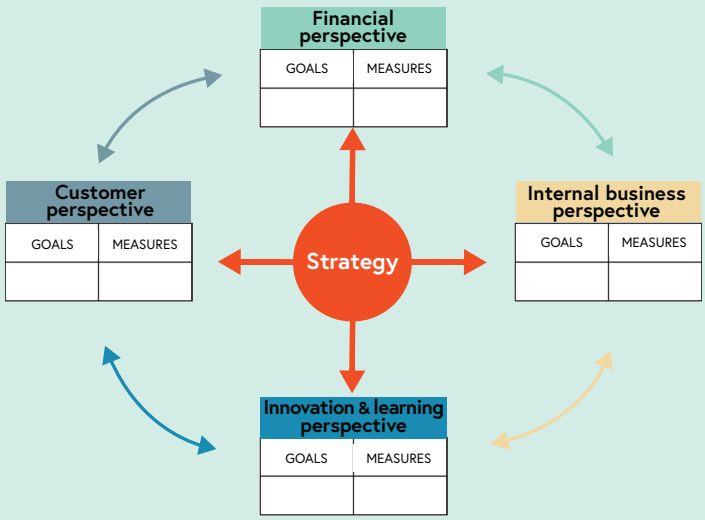


Figure 15.4 The Balanced Scorecard

The Balanced Scorecard reminds companies that financial measures are only part of the picture, but even today many companies would need this reminder time and time again.

2 Kaplan, R. S. & Norton, D. P. (1996). Strategic learning & the Balanced Scorecard. *Strategy & Leadership, 25*(5), 18–24.

Box 15.2 **The DuPont model**

The DuPont model is a schematic way to break down a company's profitability into its components. It was originally used by the DuPont Corporation in the 1920s, which gave it its name.

The DuPont model is used in finance to explain the calculations used for the financial metric *Return on Investment* (ROI). ROI is calculated by taking the profit (net gain or loss) and dividing it by the total amount invested. It is an omnipresent metric in accounting and also used to justify investments in manufacturing. Conventional wisdom suggest that 7% or greater is considered a good ROI for an investment in stocks. In specific investment decisions, the ROI percentage metric must be compared against alternative investment options. ROE is a measure of a company's efficiency at generating profits using only the shareholders' stake of equity in the business.

The point here is not to use the DuPont model to calculate ROI, but to understand the factors that ultimately drive financial results. Figure 15.5 shows the factors in the DuPont model. Note that different variants of this model has been proposed. As can be seen from the figure, it is to a dominant extent operational decisions that affect outcome ROI.

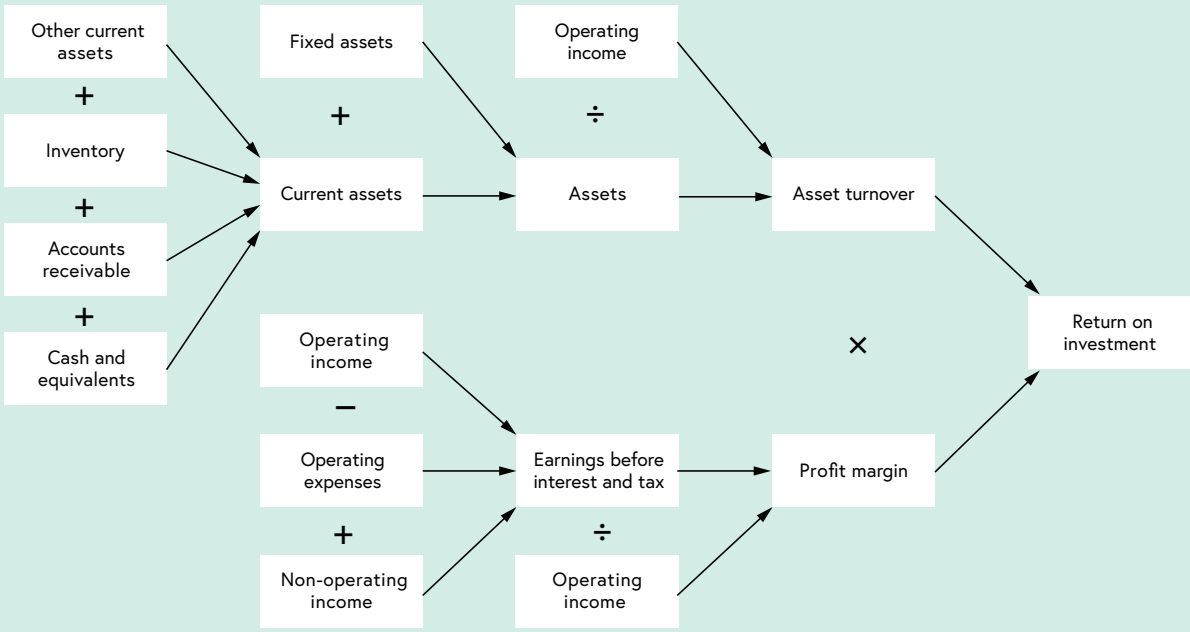


Figure 15.5 The DuPont model

The DuPont model is attractive since it is simple, but inside many of the boxes are calculations to summarize commercial transactions that are based on many assumptions. *Management Accounting* is a subject in its own right, beyond the scope of this book.

15.1.5. Management by Objectives

Management thinker Peter Drucker promoted management based on performance measurement. He called it *Management by Objectives* (MBO).[3] By today, MBO is largely debunked, at least as a superior management philosophy. W. Edwards Deming criticized it in the 1980s and Drucker himself renounced it in the 1990s. There are plenty of traps in measurement:

- The MBO aphorism "what gets measured, gets done" has some truth to it and managers risk getting exactly what they measure, not what they want. MBO also ignores the fact that many things that are not measured also get done.
- The use of performance measurement creates a measurement-driven organizational culture, not necessarily focused on customer needs.
- What is easy to measure is usually not what is right to measure.
- While measurement may create extrinsic motivation, it can hurt the more effective intrinsic motivation.
- A performance focus creates internal competition, which almost always will be gamed, and creates less cooperation and sharing between teams and departments. This is particularly a problem when performance metrics are used for compensation and rewards.
- When measures are not met, people can get low self-esteem, burnout, and low job dissatisfaction. When measures are always met, people can get complacent.
- Performance measurement systems can be costly bureaucracies that drain resources away from delivering performance.

Yet, MBO continues to be popular among many managers. More recently, it has been rephrased to *Objectives and Key Results* (OKR) and re-popularized by Intel and Google (see Box 15.3). MBO is seductive. It is simple and makes the job of "management" clear; set targets, measure, push for target achievement, report results – and repeat. As with many management ideas, the best strategy is to be pragmatic rather than dogmatic, using elements that work and discarding others.

Box 15.3 Objectives and Key Results (OKR)

Inspired by MBO, one of the founders of Intel, Andy Grove, introduced a framework for performance management, later known as Objectives and Key Results (OKR). The method is described in his book *High Output Management*[4] and later popularized by John Doerr's 2017 book *Measure What Matters: How Google, Bono, and the Gates Foundation Rock the World with OKRs*.[5] Users of OKR, including Google, credit significant and measurable improvements to the method.

While OKR is not a rigid method, it has a few important principles. One is that every objective (O) should be significant, concrete, clearly defined, and inspirational(!). Each objective should have 3–5 key results (KR) tied to it. KRs should be unambiguously measurable, in such a way that one can answer the question "Did we achieve that result? Yes or no." The target success rate for KRs should be 70% according to Doerr. OKR prefers leading indicators to lagging indicators. *Tasks* are planned to achieve the key results.

3 Drucker, P. (1954). *The practice of management*. Harper/Heinemann. (2007, Rev. ed). Butterworth-Heinemann.

4 Grove, A. S. (1995). *High output management*. Random House. (Original work published in 1983).

5 Doerr, J. (2017). *Measure what matters: How Bono, the Gates Foundation, and Google rock the world with OKRs*. Penguin Publishing Group.

Hoshin kanri (方針管理), discussed in Box 15.4, is another approach to performance management, focusing less on metrics. Hoshin means "direction" or "compass needle" and Kanri means "control" or "management." It was developed in Bridgestone in 1964, and is used by, among others, Toyota and HP since the 1980s.

As told by Isao Yoshino,[6] managers at Toyota's headquarters in Japan developed hoshin kanri and A3 thinking practices from 1980 to 1982, as part of a management development program called "Kan-Pro." Hoshin kanri is also known as "hoshin planning" or "strategy deployment."

Box 15.4 **Hoshin kanri**

While inspired by MBO, hoshin planning is focused on strategic objectives that may not be expressed in terms of metrics. In 1997, as reported by Pascal Dennis, one of CEO Fujio Cho's hoshin for Toyota in North America was "Beloved company." In terms of expectations, it's not as straightforward as "increase sales by 20%."

The engineering team's interpretation of this hoshin was to design the Prius. In Pascal Dennis's HR office in Canada, it was to organize an open-house in the factory and strengthen its ties to the community. As described by Dennis,[7] each manager's hoshins fit on an A3, and A3s have been the primary mode of communication within the hoshin kanri process.

In hoshin kanri, managers start with five or six strategic directions that may or may not be reducible to a metric. Hoshin kanri sets strategy for the next year or the next three years. It cascades the directions set by top management through the organization and translates them into actions and projects at every level. It is not, however, a top-down process but an interactive one, with a give and take within and between management layers that alters the strategic direction itself.

Known as the "catchball process," top-managers initiate an annual strategy planning process and toss it like a ball to the middle management, who respond with tactical goals and toss it back. After consensus is reached, the middle management repeats the process by communicating the goals to the front-line management, who responds with operational goals. Regardless of hierarchical level, every organizational level has a defined activity plan for the year that is aligned with the overall annual strategic goals of the organization.

At Wiremold in the 1990s, ERP rose to the top of the priority list in the 5th year of the company's lean transformation. Success had led the company to acquire competitors, and the acquisitions had their own, obsolete legacy systems. These were obstacles to the integration of all their businesses. At that point, management decided that it made sense to replace the legacy systems with a new system, and made it a Hoshin for the year.

6 Anderson, K. (2020). *Learning to lead, leading to learn: Lessons from Toyota leader Isao Yoshino on a lifetime of continuous learning*. Integrand Press.

7 Dennis, P. (2006). *Getting the right things done: A leader's guide to planning and execution*. Lean Enterprise Institute.

15.2. Metrics in manufacturing

Metrics in factories are often organized by dimensions of performance, like "safety," "quality," "delivery," "cost," "environment," and "people" (SQDCEP), but many variants exist. Very common is also "productivity." Others replace "people" with "morale," and others again use "time" instead of "delivery." The dimensions under which one sorts the metrics is less important than choice of metrics themselves.

A wide range of different types of metrics are used in manufacturing, but some are more common than others. The next sections review typical metrics for the following performance dimensions:

- Productivity
- Quality
- Equipment
- Lead time
- Inventory
- Safety
- Environment
- Cost

15.2.1. Metrics of productivity

Many companies measure productivity in terms of Sales/Employee. There is an easy way to boost this metric; outsource all production, close all plants, and become a trading company. It is not easy to find metrics for productivity that are meaningful and cannot be gamed, but it can be done. For overall company productivity, for example, you can use Value added/Employee, where

$$\text{Value-added} = \text{Sales} - (\text{Material costs} + \text{Energy costs} + \text{Outsourced Services costs})$$

This is what Peter Drucker[8] called *Contributed Value*. Materials, energy, and outsourced services are the external inputs needed for the sales. Value added/Employee is not a perfect metric, but at least it eliminates the perverse incentive to outsource, and many federal bureaus publish statistics on value-added and employment by industry that are helpful for high-level benchmarking. The ratio of value added to external inputs can be compared to another company's, or to industry averages, regardless of currency.

Because Value added/Employee is in the language of money, employees in operations cannot easily relate it to their work or see how their actions can affect it. In the language of things, you can count the number of units produced per operator per period or measure the number of labor hours per unit. It can be used to track improvement *within a shop* but not to compare factories or companies. Unfortunately, it is also a metric that hides important details, such as levels of outsourcing exemplified in Box 15.5.

8 Drucker, P. (2012). *Management*. Taylor & Francis.

Box 15.5 **Employees per car: VW vs. Toyota**

In 2015, VW employed 600,000 people to produce 10 million cars while Toyota employed 340,000 to produce just under 9 million cars. Since VW produces 10 million/600,000 = 16.67 cars/employee/year, and Toyota 9 million/340,000 = 26.47 cars/employee/year, one is tempted to conclude that Toyota is 60% more productive than VW.

If a company employs more people than another to produce the same output, yes, it may be because it is less productive. But it may also be because it goes deeper into the manufacturing process. In car manufacturing, depth ranges from modular assembly, where, at Smart in 2014, 863 employees assembled a few major subsystems built by suppliers into 87,195 cars, or 101 cars/employee/year to the old Ford River Rouge model, where, in the 1930s, 100,000 employees built finished cars from ore and sand, at about 10 cars/employee/year.

The Rouge is relevant, as it was the inspiration for the VW "mothership" in Wolfsburg, down to its smokestacks. According to a 2012 *Autoblog* article by John McElroy, VW is as vertically integrated today as GM was at its peak 50 years ago, and makes in-house many of the components that Toyota buys.

The number of cars/employee/year cannot be used to compare productivity between car companies with different manufacturing depths. Perhaps VW is lagging behind Toyota in productivity, but this metric does not prove it.

15.2.2. Metrics of quality

In the language of money, there are *Cost of Quality* (COQ) metrics. In their standard form, however, they are limited to the direct costs of failure, appraisal, and repair, which do not even begin to account for the sales impact of the company's reputation for quality. In the Quality Control Handbook, Frank Gryna[9] explicitly excluded such indirect items from the COQ as being too complex and too controversial. The problem is that when indirect costs or poor quality are excluded, the COQ metric is incomplete and can lead to serious misuse.

Stop and Think 15.2
W. Edwards Deming said, "No one knows the cost of a defective product – don't tell me you do. You know the cost of replacing it, but not the cost of a dissatisfied customer." What do you think he meant with this statement?

Many companies use a scoring system, as illustrated in Figure 15.6, to measure outgoing quality. Starting from an agreed score for a perfect unit, points are deducted for every defect observed. The Quality department publishes averages of these scores by shift, day, or month, looking for an upward trend.

9 Gryna, F. M. & Juran, J. M. (1988). *Juran's quality control handbook.* McGraw-Hill.

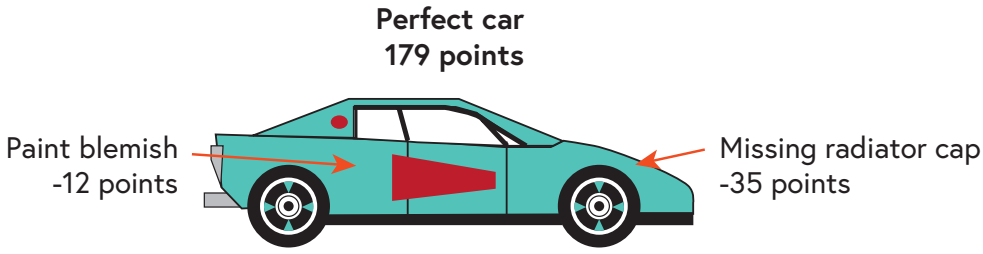


Figure 15.6 A scoring system for cars

While such metrics are meaningful to those who develop them, they are unintelligible to the rest of the organization. Higher scores are understood to be better than lower scores, but the relationship of the score to competitive position and the sources of its variations are difficult to interpret. In addition, the penalty assigned to each type of defect can become a matter of debate.

Quality is more effectively measured by using multiple, simpler metrics, covering different subtopics, such as:

- **Ratings by external agencies for consumer goods.** In the automotive industry, the JD Power and Associates Initial Quality Study rates products based on the number of problems reported by buyers within the first three months of purchase, which reflect manufacturing quality. Consumer Reports publishes reliability and maintainability ratings for cars that have been on the road for several years, which are more indicative of product design quality.
- **Counts of customer claims.** For parts sold to OEMs, Quality Problem Reports (QPR) are the equivalent of traffic tickets issued by customers. They require failure analysis by the supplier's Quality department and a report to the customer in a prescribed format, such as Ford's "8D" or Honda's "5P." The rate at which such claims are issued is clearly a metric of the customers' perception of your product quality.
- **Rejection rates.** Defectives are weeded out at the receiving process, at various points inside the manufacturing process, at final test, and at customers' receiving process. Rejection rates are calculated at each of these points and fed back to the managers in charge. There may, however, be too many of these rates to be usable by top management.
- **First-Pass Yield, also known as Right-First-Time.** The First-Pass Yield may be the most useful measure of in-process quality for top management. It is the percentage of the production volume that makes it through the entire process unscathed – that is, without failing any inspection or test, and without being waived, reworked, or scrapped (see Figure 15.7).

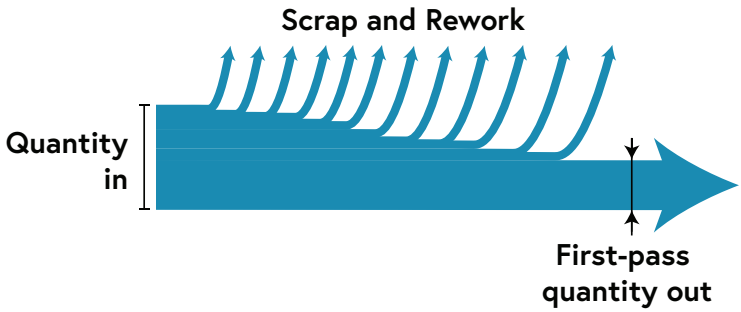


Figure 15.7 The First-Pass Yield metric

15.2.3. Equipment metrics

The most commonly (mis)used metric for equipment performance is *Overall Equipment Effectiveness* (OEE):

$$\text{OEE} = \text{Availability} \times \text{Performance} \times \text{Quality}$$

It is based on the theory of major losses that is the basis of *Total Productive Maintenance* (TPM) but it was not part of TPM as initially developed in Japan. There is no mention of OEE in the original reference on TPM by Seiichi Nakajima.[10] There is a section called "Striving for overall equipment effectiveness" but it's not about a metric called "OEE." It's about eliminating the "6 big losses" that are supposed to be aggregated now in the OEE. OEE most likely emerged outside of Japan when consultants under the influence of Peter Drucker learned about TPM. Since then, manufacturers have diverted much effort to calculating OEEs.

The 3 factors of the OEE are defined differently in different organizations. There are issues with all 3 factors in the OEE formula:

- **Availability.** The *availability* of any device is the probability that it can be used when needed, as in the probability that a spindle is up and ready whenever you have a workpiece to put on it. In the OEE context, it is usually calculated at the ratio of net time available to assign it work in a planning period to the length of this planning period. If, in a 480-minutes shift, a machine stops during a 30-minutes break and has up to 60 minutes of unscheduled downtime and setups, then the planner can count of 480 - 30 - 60 = 390 minutes in which to schedule work, which yields a ratio of:

  $$\text{Availability} = 390/480 = 87\%$$

  This assumes that the machine's ability to do work is proportional to the time it is up. For example, your connection to a server may work 99% of the time while uploading a large file and break every time you try to save it. The formula makes it look as if it has 99% availability when in fact it is 0%. This is not to say that the formula is wrong but only that it commingles the effects of many causes and that its relevance is not universal.

  There may be better ways to quantify availability depending on the characteristics of a machine and the work it is assigned. Companies that calculate OEEs often do not bother with such subtleties and simply equate availability with uptime.

- **Performance.** Performance is a generic term with many different meanings. As a factor in the OEE, it is the ratio of nominal to actual process time of the machine. If the machine actually takes 2 minutes to process a part when it is supposed to take only 1, its performance is 50%. The times used are net of setups and don't consider any quality issue, because quality is accounted for in the last factor. This factor is meant to account for microstoppages and reduced speeds, and it is a relevant and important equipment metric in its own right.

- **Quality.** Quality is not a metric but a whole dimension of manufacturing performance with many relevant metrics. In the OEE, this factor is just the *yield* of the operation, meaning the ratio of

10 Nakajima, S. (1988). *Introduction to TPM: Total Productive Maintenance*. Productivity Press.

good parts to total parts produced. It is not the *First-Pass Yield*, because reworked parts are still counted as good.

While the OEE summarizes metrics that are individually of interest, not much use can be made of it without unbundling it into its different factors. Since the meaning and the calculation methods for its factors vary across companies, it cannot be used for benchmarking. Yet, people erroneously use OEE for benchmarking all the time.

Who doesn't want all machines to be always available when needed, perform to standard, and make good parts? The problem is that, in practice, increasing the OEE is often confused with increasing utilization, and there are machines for which it is not a worthwhile goal. Such machines may be cheap auxiliaries to expensive ones, like a drill press following a large milling machine in a cell, or they may have been bought for their ability to take on a large variety of tasks on demand.

The way to use OEE is first to be very cautious about its use and select machinery that are bottlenecks or very expensive. Unbundling the OEE into its component factors yields a more easily understandable set of equipment metrics that is less likely to mislead management. While these metrics can be collected on each piece of equipment, management must then be wary of aggregating them over machines that are intended to be used differently.

15.2.4. Metrics of lead time

Lead time metrics are popular for good reasons, but there are many different lead times that are of interest and they are not easy to define, measure, or interpret. In its most general form, the lead time of an object through a system is the interval between the time it enters the system and the time it leaves it. The objects can be material, like manufacturing work pieces, or data, like a customer purchase order, or a combination of both, which starts as data and ends as materials accompanied with delivery documents. The system is any part of your business that you can draw a boundary around and monitor objects going in and coming out.

Order Fulfillment Lead Time, also known as order fulfilment cycle time, or order lead time, is, in principle, well defined as the interval between the placement of the order and receipt of the goods by the customer. The objects are orders and the system is comprised of your own company and its distribution network. There is no ambiguity as to the time the order is placed when a consumer confirms the checkout for an online cart, nor is there about the time of delivery when it is recorded by the delivery service. On the other hand, business-to-business transactions frequently do not have that clarity, particularly on large, long-term orders. If a customer places an order for 12 monthly deliveries, strictly speaking, the order fulfillment lead time is one year, which is not terribly useful. Then you have to identify a trigger point to start the clock for each delivery. If you use kanbans or other types of pull signals, they can be used for this purpose.

Inside the company, if production work is organized in jobs or workorders, you can measure the time between release by production control and completion, and that gives you an internal *Manufacturing Lead Time*, also known as production throughput time. If you produce the same item every day one piece at a time, you can record the times through a production line by serial number. But the existence of scrap and rework makes this a bit more complicated. The parts that do not make it out of the line tie up capacity and slow down the others, and the parts that are reworked undergo extra processing, adding

to the lead time and increasing its variability. When calculating lead times for a process, however, you should only consider the units that make it out as good product.

An assembly process involves multiple flows merging. It is somewhat like a river basin, and there is often no objective criterion for deciding which of two merging rivers is the main stream and which one the tributary. Usually, the smaller river is designated as the tributary, but there are exceptions. By this criterion, for example, the river crossing Paris should be called the Yonne rather than the Seine, because, as French kids learn in primary school, where they merge upstream from Paris, the Yonne is the larger of the two. Likewise, in assembly, working upstream from the end, you have to decide which flow is the main and which are feeder lines. It is a simple call where you are mounting a side rear-view mirror on a car, but less obvious when you are mating an engine and transmission assembly with a painted body.

Tracing the lead time of completed units through multiple operations requires a history database with timestamps for all the relevant boundary crossings. This is only available if there is a tracking system collecting this data. If the data collection is manual, it often occurs at the end of each shift, meaning that we know in which shift the event occurred but not at what time within that shift. To measure lead times in weeks, it is accurate enough; in hours, it isn't.

The direct measurement of lead times is also problematic with rapidly evolving, high-technology processes like semiconductor fabrication, which have manufacturing lead times in months. If a part goes through 500 operations in 4 months, its actual lead time will commingle data about current conditions at the last operation with 4-month-old data about the first one. Since then, 3 additional machines may have been brought online, 2 engineering changes to the process may have taken place, and volume may have doubled, all of which makes the old data useless.

It would be more useful to have a snapshot of the lead time under current conditions, with the understanding that it is an abstraction because, as the process keeps evolving, no actual part will ever make it from beginning to end under these exact conditions. To get such a snapshot, we need to measure lead times for individual operations, also known as process cycle times, which raises the question of how we can infer lead times for an entire process from operation lead times.

One mistake that must be avoided is to add all process times in the ERP system and use it as a standard for lead time. Process times in the ERP system are extreme values, set to their maximum to ensure processes have enough time. Adding extreme values will not represent the true lead time.

15.2.5. Using inventory metrics to calculate lead times

Inventory is easier to measure than lead time and, under stable conditions, Little's law provides some information about average lead times, and details on lead time distributions can be inferred from inventory age statistics. In addition, inventory metrics are useful in their own right, to support improvements in storage and retrieval, materials handling, and supply chain management.

Inventory levels are often expressed as *Days of Inventory*, which means days of coverage with the average production volumes. Two hundred units in stock, consumed at the rate of 10 units/day, will last 20 days. Therefore, 200 units is equivalent to 20 days of inventory, and this is what the average lead time

for 1 unit will be. This is the method most commonly used to assign durations to "non-value added activities" in Value Stream Mapping.

It is an application of Little's law. Since we can count inventory and measure throughput, we can infer average lead times from just this data. One snapshot will not give you an accurate estimate, but it is still considerably easier to take a few snapshots of a production line to get a more accurate estimate than it is to research history. The point is to get close to an answer that would take much longer to get if you actually had to be accurate.

We should not forget, however, that the validity of this number is contingent on consumption. If it doubles, the same number of parts represents 10 days instead of 20. If consumption drops to 0, then the 200 parts will cover the needs forever. When, on the basis of today's stock on hand and today's throughput, a manager declares that it is "20 days of inventory," it really means one of the two following assertions:

1. If we keep producing at the exact same rate, the current stock will be used up in 20 days, which is simple arithmetic.
2. If the production rate and available stock fluctuate around the current levels, the item's lead time from receiving into the warehouse through production will fluctuate around 20 days, by Little's law.

In either one of these interpretations, we have an "instantaneous" lead time that is an abstraction, in the sense that no actual part may take 20 days to go through this process, just as a car going 60 mph this second will not necessarily cover 60 miles in the next hour. In the case of a car, we all understand it is just a speedometer reading; for days of inventory, it is easy to draw conclusions from the number that goes beyond what it actually supports.

As we have seen, measuring lead times directly is difficult because it requires you to maintain and retrieve complete histories for units or batches of units. Inventory is easier to measure, because you only need to retrieve data about the present:

1. The inventory database is much smaller than the production history database.
2. Because inventory data are used constantly to plan, schedule, and execute production, it is readily accessible.

For similar reasons, throughput data is also easier to access than history and more accurate. As a result, with all the caveats on assumptions and range of applicability, Little's law is the easiest way to infer mean lead times in steady state.

In some cases, inventory data lets us infer more than just mean lead times. Often, the inventory database contains the date and time of arrival into the warehouse by unit, bin, or pallet. Then, for a relevant set of items, we can plot a histogram or the density of the ages of the parts in the warehouse, which, as a snapshot of its state, may look like Figure 15.8.

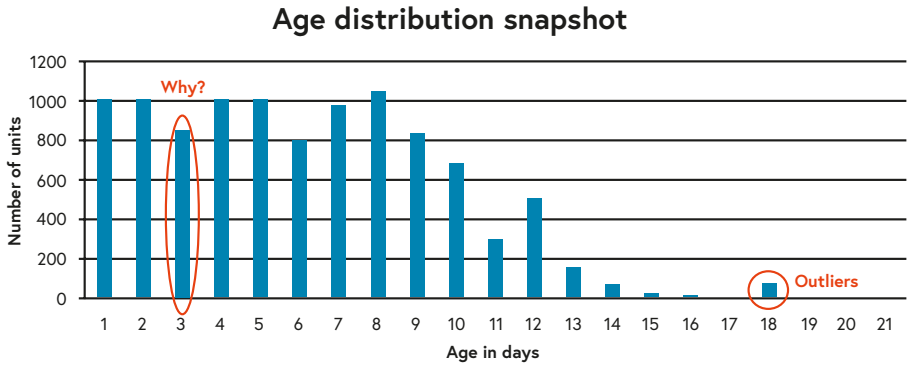


Figure 15.8 Inventory age snapshot for one item raises questions

If there is "always at least 5 days of inventory," then we can expect no part to leave the warehouse until it is at least 5 days old, and seek an explanation for the short bar at age 3 days. The bar to the right shows outliers, parts that have been passed over in retrieval for being too hard to reach, or possibly have gone through a special quality procedure. In any case, they are an anomaly that needs investigating. They may be slow-moving or obsolete goods that need attention.

If the warehouse operations are stable in the sense that there is a lead time distribution, then, if we set aside obvious outliers and take the averages of multiple snapshots taken at different times of the day, the week or the month as needed to smooth out the spikes associated with truck deliveries, the chart should converge to a pattern like that of Figure 15.9.

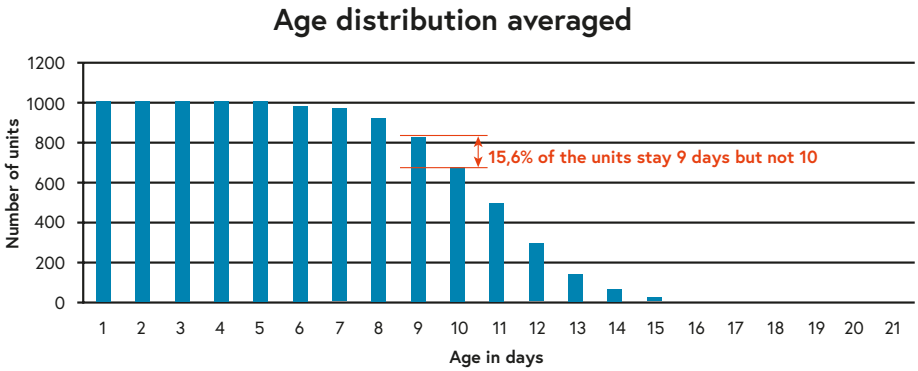


Figure 15.9 Average of multiple snapshots with outliers removed

If a unit is 9 days old in the warehouse, we know its time in the warehouse will be at least 9 days, even if it moves right away. The drop between the columns for 9 and for 10 days then represents the parts that stay at least 9 days but less than 10. In other words, in proportion to the whole, it gives the relative frequency of parts pulled on their 10^{th} day in the warehouse. Therefore, by differences, the age distribution gives us the complete distribution of the lead times, as shown in Figure 15.10.

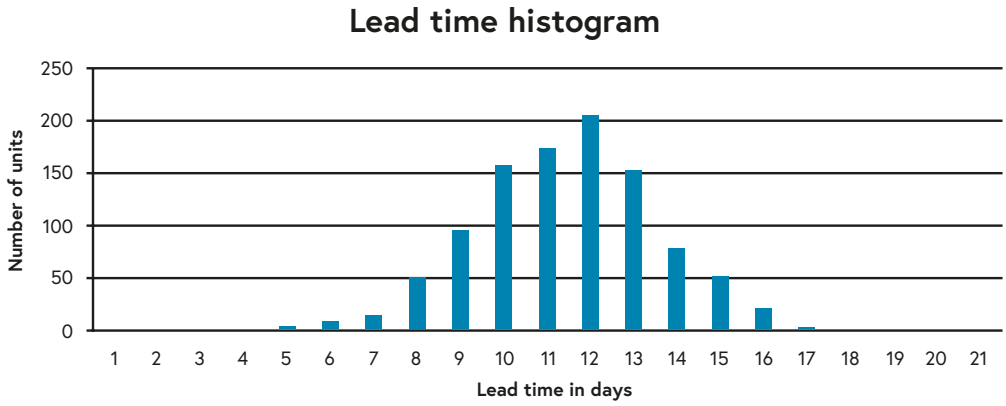


Figure 15.10 Lead time distribution inferred from inventory age

15.2.6. **Metrics of inventory**

To accountants "resource consumption" is synonymous with cost, and is expressed in the language of money. As discussed earlier, for posting on the shop floor, we need metrics that express performance in the language of things. Depending on circumstances, such substitutes may include the amount of work-in-process used to sustain production, as a measure of the effectiveness of production management and engineering. When it goes down, it is a both a one-time reduction in working capital and a reduction in recurring holding costs. The unit of measure of WIP can be set locally in each work area.

Many companies measure inventory in terms of its monetary value (*Inventory Holding Cost*), of the time it would take to consume it (*Inventory Days*), or the turnover frequency (*Inventory Turnover*). In doing so, they combine measures of the inventory itself with other parameters, such as the values assigned to inventory by Accounting and an assumed throughput rate. These are legitimate derivative metrics of interest to management but, when you stand next to a rack on the shop floor, you see pallets and bins, not money, days of supply, or turns. The raw inventory data is comprised of quantities on hand by time, and these should also be used as the basis for simple metrics in the language of things, such as the following:

- **Number of pallets, bins and units on hand by item.** This is what Operations has to work with, regardless of item cost.
- **Number of partial containers in store.** The presence of "partials" in store is a sign of a mismatch between batch sizes in quantities received and consumed.
- **Floor space occupied by inventory.** This is of interest because freed up space can be used to increase production.
- **Inventory accuracy.** This is usually measured by the percentage of items for which database records agree with reality, as observed through cycle counting.
- **Product availability.** A measure of the extent to which inventory fills its purpose of preventing shortages. Having an item in stock takes its order fulfillment lead time out of planning and scheduling. Being short an item takes order fulfilment lead time back in, with potentially disastrous consequences. We can count the shortages that do occur, and, in particular, those that cause a delay in delivery to a customer.

As discussed above, inventory is easier to measure than lead time, and much about lead time can be inferred from inventory status, using tools like Little's law or age analysis. But it is not the systematic

application of formulas to numbers: We need to be careful about underlying assumptions and the extent to which the data supports our conclusions. Accountants estimate *Inventory Holding Costs* as the sum of four elements:

1. **Lost returns.** The opportunity cost of the money invested in inventory. This money would otherwise presumably have yielded the same returns as the company's other investments.
2. **The expenses of operating warehouses.** This includes the facilities, storage devices, climate control, handling equipment, insurance, information systems, and labor.
3. **The quality cost of handling damage or deterioration.** Not only are the materials in a warehouse subject to damage from being knocked about during handling but they can also rust or spoil while on shelves.
4. **Obsolescence.** Engineering changes, regulatory changes, or the end of a product life may make materials in the warehouse unsuitable for any use, except possibly as spare parts for obsolete products.

Lacking the detailed data to put numbers on all theses categories, accountants often allow engineers to use one single ratio between 25% and 50% per year of the cost of the materials in the warehouse as the inventory holding cost in project justifications.

In practice, the amount of money tied up in inventory is often ignored in boom times but becomes a focus of management attention during busts, simply because it's not available to pay bills or wages. Then it's not a question of lost returns but of solvency. The quality component of holding costs also ignores what is often the largest contribution of inventory to quality problems: Defects go undetected while parts are in inventory.

15.2.7. Metrics of safety

Safety has become an integral part of daily measurement and reporting. Everyone understands the importance of avoiding serious injuries and fatalities but, in reality, you cannot measure a factory into a safe work environment. Complying with regulations protects management against liability and is a cost of doing business but is insufficient to guarantee safety. Accident-proofing a factory floor is similar to mistake-proofing it, and the same methods apply. It requires actively searching for accidents waiting to happen and modifying the workspace to make them impossible without adding labor.

A common lagging indicator used for safety is *Days since last injury with absence / accident*. In factories where safety is top priority, this metric is often communicated on large billboards outside the factory. It doesn't tell if the last accident was a fatality or a sprained ankle and it focuses exclusively on the last accident. It makes no difference between a factory with 5,000 employees that had an accident last week after 10 years with none and a factory with 500 employees with 5 accidents/year.

The goal of course is to show zero tolerance for accidents, and the longer a factory is able to report a perfect record, the more it hurts when the number has to be reset to 0. A key challenge in using this metric is to keep managers from pressuring employees not to report injuries and from blaming victims for hurting the company's record.

A lagging indicator is of limited use in hindering incidents from occurring. Therefore manufacturers use a lead measure to pick up trends and alert of possible future incidents: *Near misses*. The idea that near

misses can help avoid incidents comes from Herbert William Heinrich,[11] who in his job in insurance observed that in a group of 330 accidents, 300 will result in no injuries, 29 will result in minor injuries and one will result in a major injury (see Figure 15.11). This observation was popularized in Heinrich's law, which says that the number of accidents is inversely proportional to the severity of those accidents.

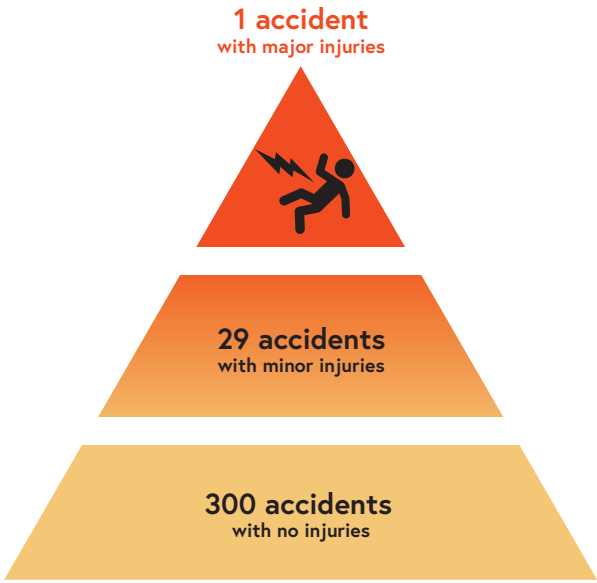


Figure 15.11 Heinrich's safety pyramid

Heinrich's law has made company track *Reported Accidents* and *Near Misses* in the hope that this will build a safety culture that avoids severe accidents. While these KPIs can be effective in their mission, it is very hard to track accurately. It is dependent on people's reporting of accidents and near misses, which requires a discipline few organizations are able to maintain. Unfortunately, the result is sometimes measurement complacency ending in a major accident, which again revives the importance of reporting and measuring near misses and minor accidents.

The safety pyramid has also been subject to criticism and called unscientific. The data and conclusions have been called into question. Some take it literally and work only on minor accidents with the hope that major accidents will then never happen. But they do.

15.2.8. Metrics of environmental performance

While sustainability was of secondary importance in manufacturing for a long time, it should not be so any more. The climate crisis has been thoroughly documented by the the Intergovernmental Panel on Climate Change (IPCC) of the United Nations and it is known that manufacturing is one of the contributors to global warming. Manufacturers – and the people who work in manufacturing – have external and personal incentives to improve this situation. In addition, governments are issuing regulations that ask companies to track and reduce their environmental footprint. To support this change, manufacturers increasingly use metrics that represent their environmental performance.

11 Heinrich, H. W. (1931). *Industrial accident prevention: A scientific approach.* McGraw-Hill.

To help identify target areas and metrics for sustainability improvement, researchers at Mälardalens University in Sweden developed the Green Performance Map[12] described in Box 15.6.

Box 15.6 **The Green Performance Map**

The Green Performance Map was developed as a response to increased awareness of sustainability in Swedish manufacturing since 2007. It is a simple framework manufacturers can use to frame their analysis of the environmental impact of their factories.

Shown in Figure 15.12, it separates environmental impact into eight areas where the left-hand side represents four inputs to manufacturing processes (value-adding materials, non-value-adding material, energy, and water) and the right-hand side represents four outputs (products, residual material, emission to air, and emission to soil and water).

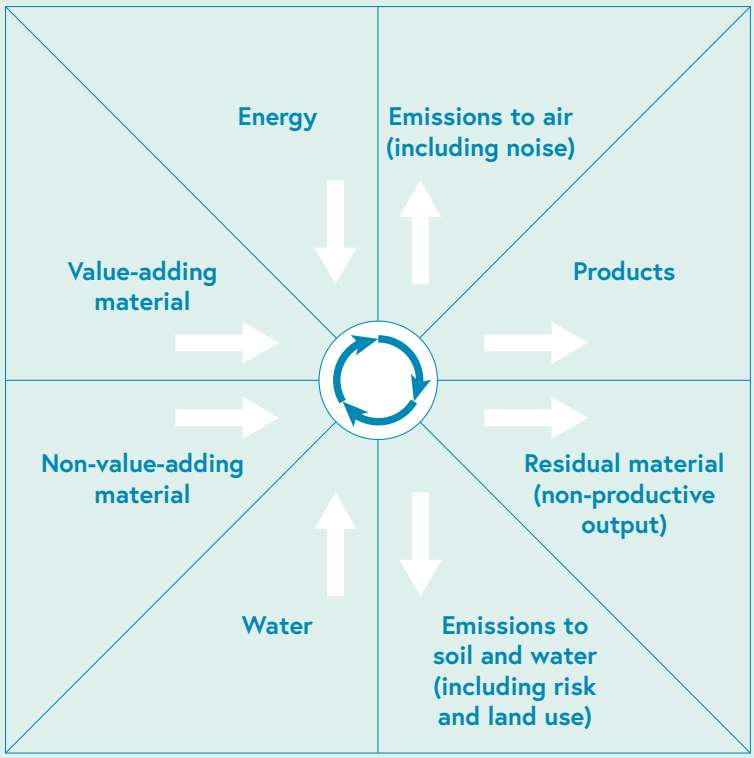


Figure 15.12 Green Performance Map by Mälardalens University

The map can be drawn on a whiteboard or printed on large-sized paper and used in workshops to guide and structure brainstorming and analysis. The map can be used at different levels, from the machine via processes to the factory.

12 Bellgran, M., Höckerdal, K., Kruvde, M., & Wiktorsson, M. (2012). *Green Performance Map handbook*. Mälardalens Högskola. / Kurvde, M. (2014). *Development of collaborative green lean production systems* (No. 162). [Doctoral Thesis, Mälardalen University]. Mälardalen University Press Dissertations, 132 pages.

Common metrics of environmental performance are as follows:

- **Energy consumption.** This is the most obvious and common metric since manufacturers also have strong financial incentives to reduce it.
- **Scrap quantity.** If manufacturers can reduce the amount of materials they scrap, they save money and reduce the environmental footprint.
- **Water consumption.** As for energy, water comes with a cost, although a much lower one.
- **Waste water disposal.** A metric that is more relevant for some than for others. The metric is particularly important in industries like textile dyeing.
- **Waste disposal.** This is waste in the sense of garbage, litter, debris, and excess materials from production processes. It does not refer to unnecessary activities as in TPS. Also for this metric, manufacturers have financial incentives to improve, because waste management comes with a price tag, especially for materials classified as special waste.
- **CO2 emission.** This metric is very hard to measure and therefore only proxied with a range of questionable calculations.
- **Percentage of energy sources coming from renewable sources.** This metric is usually only reported by companies that have the opportunity to shift to sustainable energy sources without paying a premium. Many factories have installed solar panels on the roof or even invested in their own windmills to reduce CO2 emissions.

15.2.9. Metrics of costs

One common accounting measure of costs is *Unit Cost*. A unit cost is supposed to be the total expenditure incurred to produce, store, and sell one unit of a particular product or service. Unit costs are aggregated into the *Cost of Goods Sold* (COGS) in operating statement. This metric looks simple but is based on questionable assumptions and has led many managers astray.

In everyday life, the cost of a carton of milk is the price you pay for it, and that's the starting point. If you choose not to buy it, you keep the money and it doesn't increase what you pay for anything else. In manufacturing, if you decide *not* to make a unit of product, you do not get to keep its cost and, in addition, you increase the costs of all other units you make because each one has to carry more of the factory's overhead. Some managers behave as if unit costs were like the cost of a carton of milk, as reflected in comments like "when we ship one of these, we ship a dollar with it."

Shared resources is the issue that creates problems for unit cost calculation. If you can break down your plant into focused factories, each with all the resources needed to make one product, you have eliminated the problem. Most factories, however, have not found a way to completely dedicate machines by product. Some of the support services usually stay centralized, including HR, Shipping and Receiving, IT, Technical Data Management, Safety, and others. In production, some processes still operate as common services, like Heat Treatment, Electroplating, or Painting.

Among products in a factory, only the *runners* get dedicated lines; *repeaters*, dedicated lines by product family; *strangers*, a shared job-shop. As a consequence, the problem of determining how much of a shared resource to impute to each product does not go away. In the accountant's standard answer that cost is "materials, labor, and overhead," the most problematic term is overhead, which is an allocation to each unit of product of a lump sum including everything that is neither labor nor materials. It is also the money that would have to be spent even if the factory produced nothing.

If Accounting help you administer a "black box" that puts out unit costs, it is easy to use them to choose whether to make a product or which machines/lines to use when making it. All the complexity has been put in the black box where you don't need to worry about it. This may have been reasonable in the 19th century but not today, because everybody has the computing power and standard spreadsheet software needed to do a serious analysis of the funds flows associated with a decision or event.

The concept of unit cost, while intuitive, is a mismatch to manufacturing operations. You don't make and sell one unit but a flow of units over time and this is what you should be looking at. If you increase production from 20,000 to 50,000 units/month for the next 4 years, you may need new machines and more people, materials, outsourced services, and you expect an increase in sales. In financial terms, it doesn't translate to one number but to a *schedule* of numbers over time, and it is this schedule that you use to evaluate the economics of decisions.

Generating schedules for both outflows and inflows is straightforward. It becomes complex when you apply a discount rate to the numbers for the future in order to make them comparable to the present and calculate various measures of performance. This is why it's called *Discounted Cash Flow Analysis* (DCF) and it yields different results depending on the choice of the discount rate. One commonly used method is to find the discount rate that equalizes the net present values of the schedules of in- and outflows. It's called the *Internal Rate of Return* (IRR function in Excel) and is commonly used in evaluating projects.

Because of problems with calculating costs accurately, one option is to not calculate them at all, or at least not use them for decisions on the shop floor. The case against calculating costs is that extending the concept from a carton of milk in a grocery store to a manufactured product in a factory is anything but straightforward. The results take so much energy to produce that the organization has none left to work on improvement. One consequence is that it makes you focus on the easily quantifiable and dismiss the "intangible," even when it dwarfs the tangible.

The case for calculating costs is that, unless you do, you may spend your resources on "improvements" that make no difference. For example, if you have been implementing 5S on the shop floor, you might want to apply it to the overcrowded hard disk of your server. This would mean reviewing thousands of files to decide whether they belong on the server or should be archived elsewhere. Then you might want to reorganize the folder structure and your databases and set up procedures to maintain the new organization. Alternatively, it may be faster and cheaper to just buy a bigger drive, and you need to know whether that's the case.

The problem of calculating costs "accurately" in manufacturing still persists. There is no universal solution to this problem. In an attempt to solve it, Robert Kaplan developed *Activity-Based Costing* (ABC). ABC seek to allocate the accurate share of overhead that goes into a product. To do so, it maps all the *activities* needed to produce a product, and assign a share of this activity to each product produced. Although it is complicated to administer, ABC is reportedly used in many manufacturing companies and – when it works – it is superior to simple product cost metrics. However, ABC has its own problems as exemplified in Box 15.7.

Numbers[13] Davis[14]

Box 15.7 **Activity-Based Costing (ABC) and shared resources**

Harvard's Robert Kaplan has proposed Activity-Based Costing (ABC) as an approach in which each component of overhead is allocated to products based on its own cost-driver.

ABC requires considerably more data collection and analysis. In *Real Numbers,*[13] Orry Fiume and Jean Cunningham also point out the risks of using cost drivers like the number of set-ups on a machine that may encourage line managers to do fewer setups and increase the size of production runs.

While it is complex to apply, it is still a simplistic assumption that every activity can be assigned a single driver as a measure of the amount required by a product. Imagine, for example, an electroplating facility for work pieces the size of bedsheets, in which you load the racks with parts the size of teaspoons in the gaps between the bedsheets, as in Figure 15.13. Once the racks are loaded, the direct labor to run the electroplating tanks is not affected by the presence of the teaspoons, and the extra amount of chemicals needed is imperceptibly small.

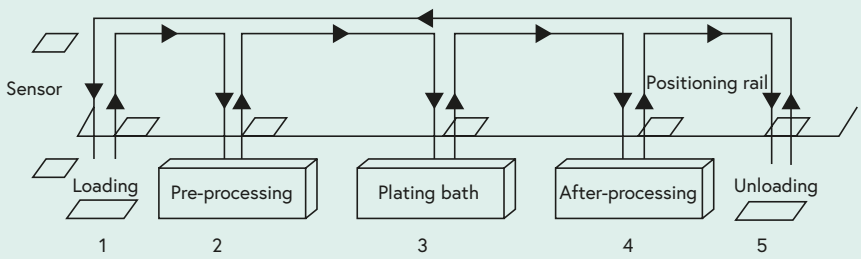


Figure 15.13 An electroplating process

How should you split the cost of electroplating between the bedsheets and the teaspoons? The Teaspoons manager thinks that piggybacking these parts on the bedsheet operation was a brilliant idea that saved the company the cost of a benchtop electroplating machine. Since it adds no expense to the factory's operations, this manager thinks it should be free. The Bedsheet manager disagrees and doesn't want to let the Teaspoons manager freeload on the racks. From this manager's perspective, the Teaspoons department should be charged the same as if it had its own machine, and that amount should be deducted from the Bedsheets' plating costs.

There is no obvious cost driver that could be used here. Morton Davis[14] addresses this question in the case of an airport runway shared by four classes of aircraft. A jumbo jet requires a longer runway than a regional jet but regional jets can use the long runway. In the simple case of the bedsheets and the teaspoons, his solution boils down to splitting the difference and charging the Teaspoons department half of what it would have spent to buy and operate its own, dedicated equipment.

13 Fiume, O. J. & Cunningham, J. E. (2017). *Real numbers: Management accounting in a lean organization*. Jean Cunningham Consulting.

14 Davis, M. D. (2012). *Game theory: A nontechnical introduction* (p. 209). Dover Publications.

15.3. **Performance boards and meetings**

Factories use boards to communicate performance in manufacturing. Students who enter a leading manufacturing facility may expect to see high-tech communication technologies everywhere – only to be disappointed when they find mostly paper-based boards or handwritten white boards on the shop floor, as in Figure 15.14.



Figure 15.14 Performance boards in factories in Poland (left) and Sweden (right)

In the era of smartphones, tablets, and liveboards, there are good reasons to be disappointed and many manufacturers are updating communications technologies for the factory floor. However, for many reasons, it is not as simple to implement and maintain as the students may think, and, more importantly, there is a value to manually updated boards that may go lost when everything becomes electronic.

The primary objective of performance boards is to make current performance and performance gaps immediately visible to the people who can influence them. A secondary objective is to collect and summarize performance for higher-level management, who may need this information to do their job or provide it to other stakeholders. Performance boards are often organized according to the aforementioned performance areas, such as SQDCEP shown on the left board in Figure 15.14.

One issue is what technology is used to create the boards; it can be a manual whiteboards, a printed piece of paper, or a fully connected digital screen. The choice of technology affects the completeness, visibility, and timeliness of performance data that can be shown, and the efficiency in terms of data collection and processing. Another issue, which is more important, is *how* this information is shared, managed, and used.

In lean production, performance boards are present in each team's work area and are quickly reviewed in a stand-up meeting at the start of every shift – a work routine also known as *Shop Floor Management*. The meetings last no longer than 5–15 minutes. Information from meetings are cascaded when needed to the next management level. While any factory can put up performance boards, be it paper-based or digital screens, few are able to build the culture needed to work routinely every day to solve problems and improve the metrics based on information from the boards and effective daily meetings.

15.3.1. **Performance board design**

The design of a performance board, in terms of content, is independent of the technology used to generate and update it. Figure 15.15 shows an effective performance board template.

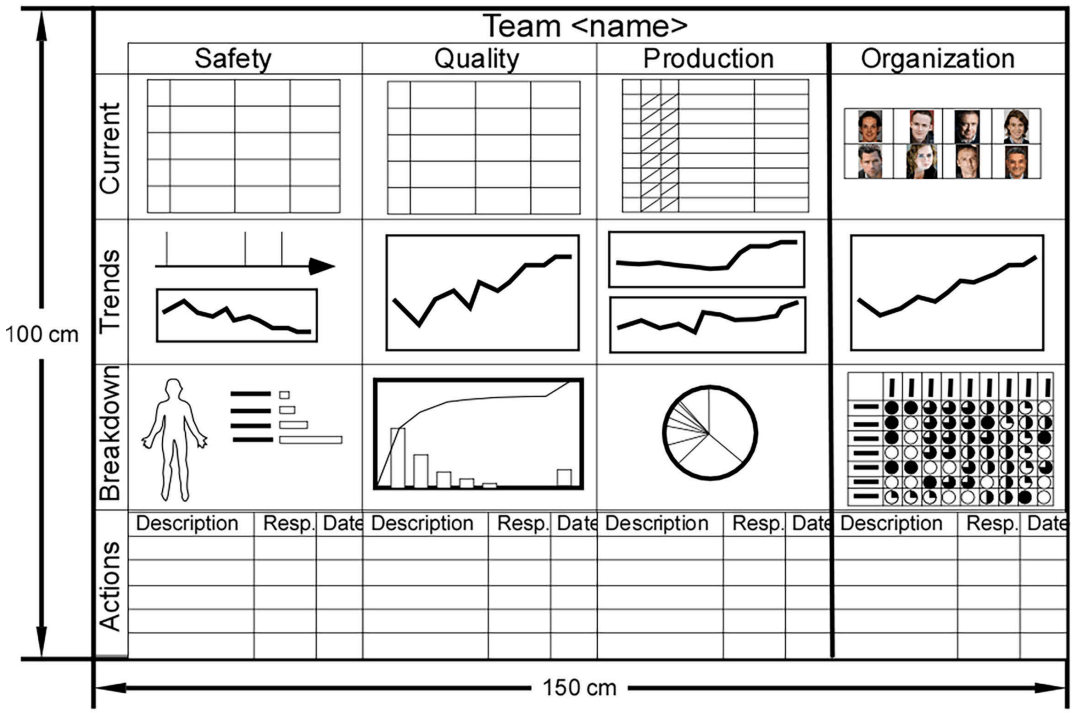


Figure 15.15 A performance board design

This template has 1 column for 3 selected dimensions of performance and 1 row for each type of informations:

1. **The top row** is for the latest news, what happened this shift or last.
2. **The second row** is for historical trends in aggregate performance.
3. **The third row** is for a breakdown of the aggregate into its constituent parts, such as the most common injuries, defect categories, most frequently made products, or the employee skills matrix.
4. **The bottom** is about actions or improvement projects in progress in each of these areas.

Some companies use a simple rule-of-thumb as best practice for board design: The 1-3-10 rule. This rule suggests that within 1 second, one should get an understanding of the purpose of the board. Within 3 seconds, one should have a rough understanding of the status: Is it going well or are there problems? Color coding with traffic light logic helps here. Is it getting better or worse? Trend charts can tell. Finally, within 10 seconds, one should have an idea what actions are taken to solve problems.

> **Stop and Think 15.3**
> Why do you think so much attention is paid to visualization when it comes to performance board design?

15.3.2. **Daily management**

Daily management meetings are central to manufacturing. They occur at different levels of the production organization and for different purposes. The number and the organization of daily meetings may vary depending on the factory's size and its structure. Most factories have a *Production Meeting* for first-line managers but the "lean daily management system" involves additional meetings for the team leaders reporting to a first-line manager and within each shop floor team.

The daily meeting are cascaded so that issues can be raised quickly and everyone gets a quick understanding of the situations that concern them, every day and early in the first shift. Figure 15.16 shows an example of possible timings and participants in a factory with three main management levels.

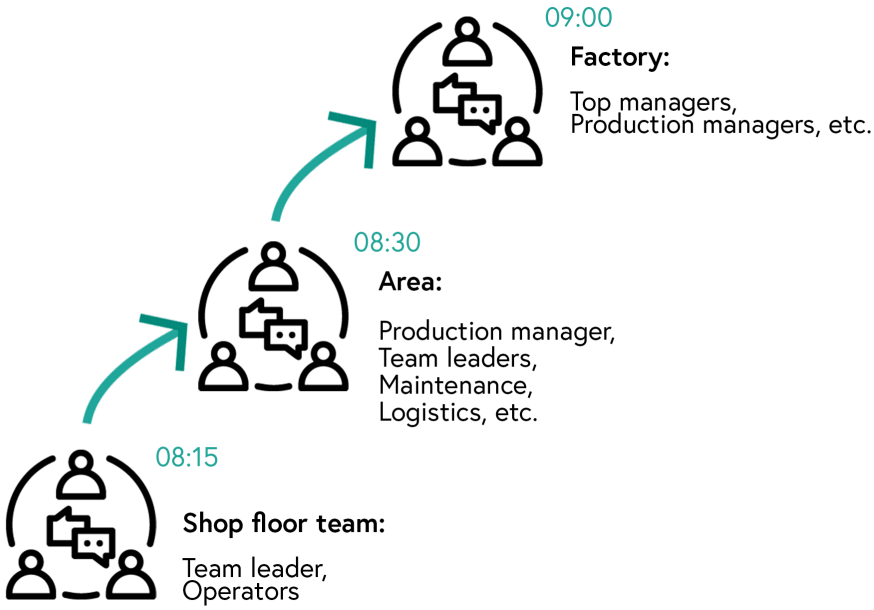


Figure 15.16 Example of daily management cascading

Shop floor team meeting

In factories that do not subscribe to lean production, it is rare to find shop floor employees engaged in daily meetings. In these factories, shop floor workers often only take job tasks delegated and communicated to them via production orders, shift meetings, or from their first line leader.

In factories that use lean production, each operator team meets in front of its performance board to review the events of the previous day, plan for any issue needing special attention today. The objective is to give the whole team transparency into the team's deliveries and performance, and raise issues directly from the front-line employees who know the situation on the shop floor best. The first-line managers may also bring all their teams together to communicate and review any announcements from management. The meetings are short, typically 5-15 minutes, and everyone stands.

Table 15.1 summarizes the main tasks and activities of production team daily meetings.

Table 15.1 Production team daily meetings

Who	What	When	How	Where
Team leader	Update board	Before meeting	Use data from prior shift and input from first-line manager	At the board
Team leader + members	Get together	Start of shift	Stand in half circle	At the board
Team leader	Brief the team: • Safety • Quality • Delivery • Productivity • Environment • Attendance • Improvements (or similar)	5 min. at start of shift	Focus on changes and discrepancies Ask for unraised issues/challenges Recognize individuals	At the board

The daily production meeting (asaichi)

The daily production meeting is a day-to-day meeting of managers from various areas, who meet to share key information, mainly focusing on deviations from the plan. It is called the asaichi meeting in Toyota. The Japanese word asaichi (朝市) means "first thing in the morning" or "morning market." It corresponds to the middle "area" layer shown in Figure 15.16.

The asaichi takes place in a specially created location at the same time every day and lasts a maximum of 30 minutes. The production meetings are often held standing and on the shop floor rather than in conference rooms. Not only does it make the meetings shorter but it also makes it easier to move participants to where the data is displayed or to actual work locations as needed.



Figure 15.17 A daily production meeting board

Daily management meeting

The daily management meeting involves the top layers of the plant management's structure:

- The plant manager
- The production managers in charge of all the lines
- The managers of the support departments: Accounting, Maintenance, HR, Supply Chain, etc.

Production managers who wish to bring up issues with top management can do it at this meeting. This meeting is also supported by performance charts at the corresponding level of aggregation. As in every meeting, the focus is on deviations from the plan. Problems should be raised so they can quickly be solved. At the end of this meeting – in factories that follow daily management vigorously – everyone in the factory are aware of the situation that affect the performance of their responsibility area. Daily management is therefore also called "daily layered accountability."[15]

While the benefits are obvious, it is often more tempting for top managers to not engage in everyday hassle, and instead call a weekly factory status meeting on Friday afternoon. Since weekly meetings necessarily must cover many more things that have happened during the week, the meetings are organized in a meeting room, with chairs and refreshments. Since they are held on Friday, the issues are dealt with at the earliest the week after.

In a lean factory, every manager up to and including the plant management should know the status of operations on a daily basis. Hence, both front-line and middle management are involved in the daily meetings. Then, issues can be escalated up the hierarchy efficiently, and information shared quickly throughout the organization. The daily meetings break down silos between departments and ceilings between hierarchical levels.

15.3.3. Manual or electronic performance boards?

Many companies have started to replace manual, analog boards with electronic boards, but it is a shift that is controversial and comes with unclear benefits. On the one side, humans pay more attention to charts drawn by the hand of someone they know than to the output of a computer system, and they are more comfortable annotating it. The hands-on manual approach empowers and engages the workforce in daily shop floor management. When information is presented on a screen, operators and managers become "blind" to it. Manual boards also allow quick experimentation and changes of board structure and content.

On the other side, to a generation that has grown up with video games and smartphones, manual boards are quaint and obsolete. Professionals that have grown up playing *World of Warcraft* will not eagerly or competently do manual data collection and analysis, but they have other skills. Boards that actually contain printouts from computer systems are not truly manual anyway. They consume considerable resources to update, by which time the data on the charts is old. Handwritten notes are unreadable, and there is no opportunity to analyze historical trends. Electronic boards can mirror the same real-time information on many boards and let teams share action points with other teams immediately.

15 Mann, D. (2014). *Creating a lean culture: Tools to sustain lean conversions*. CRC Press.



Figure 15.18 Digital board from iObeya (Source: iObeya)

While the debate is healthy, the *trend* is towards electronic performance boards, and liveboard technology makes them interactive, allowing annotation of charts with pen-like styluses. The COVID-19 pandemic accelerated this trend, as many factory managers were trying to manage their factories from home offices. It requires, however, automatic data acquisition, summarization into meaningful information, and presentation on the board. The complexity of this process is not to be underestimated.

It requires expensive investments in infrastructure and continuous upgrades by IT specialists. There are factories that have implemented expensive digital boards, only to remove them when they did not work as intended. Other factories experience log-in problems, connectivity issues, or power issues resulting in black screens that carry no information at all.

In reality, it isn't either-or; it can be a mix. Just because a factory uses automatic data acquisition and generates charts on computers doesn't mean it can't post them in hardcopy. It spares the factory management the expenses of large screens for every team, and it supports manual annotation. It also requires you to print and distribute updated charts daily, which means one day's manual annotations will be lost the following day unless more work is used on preserving them.

Annotations are not a minor issue. Spotless charts on a factory floor are probably charts that are not being used. If a team uses charts to identify and solve problems its members mark up the charts in doing so. Solving the problems is the main task, but playing it back later in one-point lessons is also desirable. For this, you need to keep the marked-up charts and preferably not on paper. With a manual board, taking pictures with smartphones to record annotations is a simple solution, as long as the plant has a system in place to organize these pictures for later retrieval.

Whatever technology is used, the more important driver of performance is how the boards are used. The choice of information to post and the way it is presented affects its usability. Whether the information is manual paper drawings or computer-generated charts, we have to learn to make it effective.

15.3.4. **Use of A3s for paper charts**

The individual charts on performance boards are often printed on A4 paper, with text that operators can't read from two feet away. A3 paper, which is twice the size of A4, with an 18pt font, is more appropriate for this function. Toyota's maintenance group in Onnaing, France, calls its performance board an "A0," which is a panel of 8 A3s.

A3s are used in operator instructions, problem solving, continuous improvement, and strategy deployment applications that have little in common beyond the value of presenting the key information in a structured, single-page format that is large enough for sharing as needed, yet supported by commonly available desktop publishing software, compatible with many copiers and printers, and with paper available as routine office supplies. Pascal Dennis[16] has described the use of A3s as central to strategy deployment (hoshin kanri) and John Shook[17] sees A3 as a "management process." Shook defines the A3 report as follows.

> **Definition**:
> The A3 report is a Toyota-pioneered practice of getting the problem, the analysis, the corrective actions, and the action plan down on a single sheet of large (A3) paper, often with the use of graphics.

Many discussions of A3 reports in TPS omit one basic fact: A3 is a *paper format*. In millimeters, an A3 sheet is 297×420 mm, roughly equivalent to 11×17 inches. It is the size of two A4 sheets side-by-side, and half of an A2 sheet. The *A-series* of paper sizes is used all over the world, except in North America, Colombia, Venezuela, and Chile in South America, and the Philippines, all of which use the US standards.

Realizing that A3 was the largest format it could run through its common office printers, Toyota made it a standard for single-page reports, work instructions, strategic plans, and other communications. An *A3 report* is not just a story on one sheet of paper, but on one sheet of paper of this particular size, which has been found optimal to tell a manufacturing story with just enough details to be useful. A3 or 11×17 is the goldilocks size for one-page management story telling. It can be posted on bulletin boards, in "Obeyas" – rooms dedicated to performance boards and problem solving –, or above operator workstations. Operator instruction sheets are supposed to be on A3 paper.

Sheet size has to do with the amount of information a reader can capture at a glance, what a group of people can see from a distance of a few feet rather than a few inches, and how easily a document can be moved, posted, or annotated. If you shrink an A3 to A4 or letter size, it no longer works as an A3, because the print will be too small for viewing on a board. If you show it on a PowerPoint slide, it is not an A3 either, because it does not have the permanence of hardcopy and you cannot easily annotate it manually.

16 Dennis, P. (2006). *Getting the right things done: A leader's guide to planning and execution.* Lean Enterprise Institute.

17 Shook, J. (2008). *Managing to learn: Using the A3 management process to solve problems, gain agreement, mentor and lead.* Lean Enterprise Institute.

15.3.5. Metrics gaming and how to prevent it

As practiced in many companies, Management by Objectives (MBO) imposes numerical targets on a few metrics, cascading this approach down the organization and giving individuals a strong incentive to improve their numbers. It is a caricature of the process Peter Drucker recommended half a century ago. The result is illustrated in Figure 15.9; KPIs looks great, but there is no effect on the bottom line.

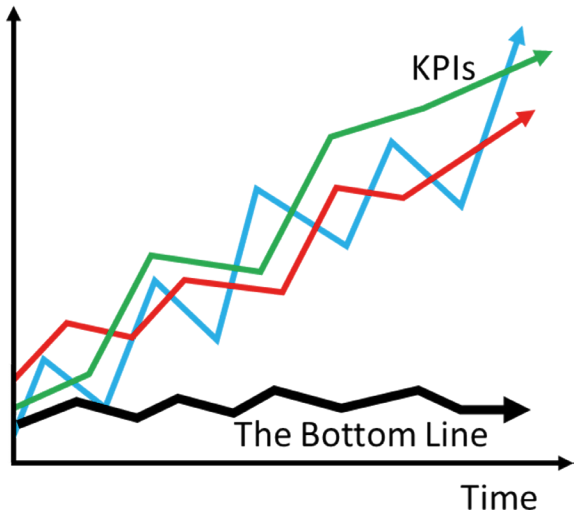


Figure 15.19 Metric gaming is a rational but non-value-adding behavior (Credit: Christoph Roser)

Recognizing metric gaming

Some of the most common ways of gaming metrics are as follows:

- **Cherry-picking.** Since there is often an abundance of metrics to chose from, the easiest way to look good is to showcase selected metrics and hide others.
- **Taking advantage of bad metrics.** Many metrics commonly used in manufacturing are poorly defined, providing gaming opportunities. For example, "sales per employee" can be increased by outsourcing activities, which reduces headcount.
- **Stealing from the future.** In sports, nothing is more dramatic than the game won by points scored in the last seconds of a game. In business, the end of an accounting period is the end of a game, and, as it approaches, sales scrambles to close last-minute deals and manufacturing to ship a few more orders. This is what Eli Goldratt called the "hockey stick effect." This is done by moving up activities that would otherwise have taken place a few days later, during the beginning of the next accounting period. As a consequence, the beginning of the period is quiescent. Not much is going on, but it is made up at the end.
- **Redefining 100%.** Many ratios, by definition, top out at 100%. A machine cannot run 25 hours/day, and a manufacturing process cannot produce more good parts than the total it makes. This is why ratios like equipment uptime and first-pass yield top out at 100%. Any result under 100%, however, invites questions on how it could be improved. A common way to fob off the questioners is to decree, for example, that a particular machine could not possibly be up more than 85% of the time, and redefine the scale so that 85% uptime is 100% performance. For production rates in manual operations, the ratio of an operator's output to a work standard is often used instead of process times or piece rates. Such ratios have the advantage of being comparable across operations,

and are not bounded in either direction. But their relevance depends on a work standard, and, when everybody in a shop performs at 140% of standard, chances are that the standards are engineered for this purpose.

- **Leveraging ambiguity.** Terms like availability, cycle time, or value added are used with different meanings in different organizations, creating many opportunities to game the metrics. If the product's market share in the first quarter went for 1% to 2%, it doubled, but, if it went back to 1% in the second quarter, it went down by 1%.

It is not that gaming is human nature but, in organizations that use MBO, organizational culture will eventually weed out those who don't engage in it. Changing such habits in an organization is not easy.

The metrics validate the approaches taken to reach the goal, but the goal is not to improve the metrics. It is a subtle but very important difference. If you have the flu, you have a fever, but your goal is to heal, not just to bring down the fever. Once you are healed, your fever will be gone, and the decrease in your temperature is therefore a relevant indicator of your healing process, but it is not the healing process. If bringing down the fever were the goal, you could "game" your temperature and bring it down without healing. This distinction existed in Drucker's original writings about MBO, but got lost in implementation.

Preventing metric gaming

What can be done to prevent metrics gaming? Below are effective tactics to reduce it.

1. **Stamp out fear.** This was the 8th of Deming's 14 points to management[18] but he was short on how to do it. A key starting point is to decouple metrics from an individual's compensation or career opportunities. Solving problems rather than searching for guilty parties to punish gradually builds a culture where people are not afraid of reporting poor performance, embrace opportunities to improve, and call for help as needed.
2. **Review the metrics themselves.** You may not be able to completely game-proof your metrics, but at least you could make sure that they make sense for your business and are not trivially easy to game.
3. **Focus on improvement, not benchmarking.** Since metrics across lines or factories never have the exact same internal and external conditions, there is little point in comparing them. Instead, focus on each team improving the performance in terms of its own metrics. While improvement can also be gamed for a short period, it is almost impossible over time.
4. **Eliminate piece rates.** Piece rates used to be the most common form of payment for production work, but have almost entirely vanished in advanced manufacturing economies, and been replaced by hourly wages. Performance expectations are attached, but there is no direct link to the amount produced in a given hour of a given day. There are many reasons for this evolution:
   - The pace of work is often set by machines or by a moving line, rather than by the individual.
   - The best performance for the plant is not necessarily achieved by every individual maximizing output at all times.
   - More is expected of all individuals than just putting out work pieces, including training or participating in improvement activities.
5. **Increase the measurement frequency.** The longer the reporting period, the more opportunities it

18 Deming, W. E. (1986). *Out of the crisis: Quality, productivity and competitive position.* Cambridge University Press.

offers for gaming the metrics by stealing from the future, and the more pronounced the hockey stick effect. Conversely, you can reduce it by measuring more often, and eliminate it by monitoring continuously, as is done, for example, by the electronic production monitors that keep a running tally of planned versus actual production in a line during a shift. Periods exist in accounting because of the limitations of data processing technology at the time the accounting methods were developed. In the days of J.P. Morgan, *closing the books* was a major effort that a company could only do every so often. In the 21st century, there is no technical impediment to the "anytime close," but the publication of periodic reports continues by force of habit. Metrics in the language of things as well as the language of money can be monitored continuously.

6 **Have third parties calculate the metrics.** In principle, counting chips should be done more accurately by agents with no stake in where they may fall. In practice, it is not only expensive but does not always produce the desired result. It is the approach used in Management Accounting. A plant's accounting manager, or comptroller, is not chosen by the plant manager, he or she reports directly to corporate finance, and has no motivation to humor the plant manager. This is a double-edged sword because, with neutrality, comes a distance from the object of the measurement that may cause misunderstandings, and Management Accounting leaders have been struggling with the challenge of providing relevant, actionable information to managers for the past 30 years.

7 **Automate data collection and metric calculation.** For metrics in the language of things, the closest to having a third party produce the measurements is to have a computer system do it, based on automatic data acquisition. This reduces the opportunity for gaming, but the issues of relevance are as acute as in Management Accounting.

8 **Reduce the dependency on metrics.** There are alternatives to management by objectives. For example, managers can manage by action and values.

15.4. **Responding to emergencies**

Emergencies are inevitable. Sooner or later, natural or human-made disasters will disrupt manufacturing operations. Most of this book discuss manufacturing during *normal* operations with activities subject to fluctuations around smooth trends. For events that are *expected*, you have pre-planned responses, like enough materials in-house to keep trucks delayed by traffic from stopping production or team leaders who can step in while an operator temporarily leaves the line. For events that are *unexpected*, you have to activate emergency plans or improvise. Emergencies test the resilience of any organization. Those who respond well come out strengthened relative to those who don't.

Manufacturing facilities and supply chains are disrupted by fires, earthquakes, tsunamis, floods, pandemics, strikes, wars, and other natural or human-made disasters. In recent years, emergencies have almost become the rule rather than the exception. Events like the US–China trade war (2018–), Brexit (2020), the COVID-19 pandemic (2019–), the global chip shortage (2020–), the six-day blockage of the Suez Canal caused by the Ever Given container ship (2021), the 2022 Russian invasion of Ukraine, cyber attacks, and many other small and large events have taken manufacturing companies from one crisis to the other. Add to this the effects of global climate change and global warming.

If the designers of the Fukushima nuclear power plant in Japan had considered the possibility of a 14m high (40ft) tsunami, they would have located the emergency generators and coolant pumps higher. They didn't, and the 2011 earthquake with a magnitude of 7.0 and the resulting tsunami caused a

meltdown as a result. It also caused a major disruption in the Japanese automotive supply chain, forcing Toyota to close 26 of 30 production lines in Japan. How do you react once it has happened?

Risk management is about identifying, setting priorities, mitigating risks, and, when all else fails, responding to emergencies. It is an essential skill in an ever-changing and global business environment. To better prepared organizations, crises are actually opportunities to seize markets or take over competitors. Risk management fills volumes of textbooks on its own. However, emergencies are too unpredictable in type and scope to have planned responses ready for all types. Besides evoking emergency plans there is no choice but to improvise an ad-hoc response. There is not, and cannot be, a general theory of how to do it. What we can say is that not all organizations are equally adept at improvising countermeasures to what can be existential threats, and that it is a skill that can be developed.

Companies that perform above average in times of crises usually focus on the following tactics:

- Use scenario planning.
- Consider a dual source strategy for critical parts and products.
- Increase the safety stock of critical parts, but not of all parts.
- Increase supply chain vigilance.
- Increase supply chain resilience.
- Build fire-fighting capabilities but don't worship them.
- Use manufacturing repurposing to help others or take a business opportunity.

These tactics are discussed below, drawing on several examples from Toyota. Also in this area, Toyota has demonstrated that their production system has specific advantages.

15.4.1. Use scenario planning

How do you run your operations if the current business is interrupted? While it is impossible to answer this question for all kinds of future events, asking it is an important first step towards increased resilience. It begs the question, "What kind of interruption?" Companies should have an idea of what could happen, when, where, how likely it is, and what the impact could be.

Not all, but many interruptions are announced as likely to happen before they actually happen. They range from early indicators to extreme weather events, wars, price shifts and capacity issues in shipping, and changes in tariffs, trade agreements, law, and taxes. For such events, companies usually have a short window of opportunity to prepare a response.

Scenario planning is a way to take control of uncertainty by identifying assumptions about the future and determining how your organization will respond to them. Scenarios are stories of alternate futures. There are many different ways and suggested road maps on how to do scenario planning, but they always include fundamental steps like the following:

1. Identify possible future events and scenarios.
2. Answer "what happens if" questions for each scenario.
3. Plan countermeasures for the most likely and critical scenarios.

An example of scenario planning is given in Box 15.8, which discusses NUMMI's proactive response to the Mississippi flood in 1993. To avoid their factory in California being affected by flooded rail transportation lines, they pre-booked road transportation capacity ahead of the flood.

Box 15.8 **NUMMI's response to the Mississippi flood 1993**

From 1983 to 2009, Toyota operated an assembly plant in Fremont, CA, as a joint-venture with GM called "New United Motor Manufacturing, Inc.," or NUMMI, which received many parts from a cluster of suppliers in the Chicago area. At a consolidation center in Chicago, these parts were sequenced and loaded onto intermodal containers that rode trains to California and were transferred to trucks for delivery to NUMMI.

In July of 1993, the Mississippi River flooded, putting the Union Pacific railroad tracks underwater. NUMMI received some 55+ Trailer on Flat Car (TOFC) per day from Chicago, and those trailers could no longer get through, potentially shutting down the plant. NUMMI watched the weather reports very carefully. The flooded area is shown in Figure 15.20.



Figure 15.20 Mississippi flood of 1993

Three days prior to the flood, as the news reports were that a flood would be severe, but nobody was saying for sure that it would happen, NUMMI took preemptive action and got commitments from all the trucking companies in the Chicago area. They were hedging supply transportation options. Had the flood not materialized, they would have lost thousands of dollars in "trailer holding charges."

When the flood actually came, NUMMI was prepared and nobody else was. The cost of the NUMMI assembly line being down for one hour was about $750,000. The incremental daily cost of shipping 55 trailers over the road, as opposed to by train, was about $75,000. This is the equivalent of 6 minutes of lost production, and no questions were asked.

After the flood receded, NUMMI continued with trailers for a few days until the congestion had cleared on the railroad, and then switched back to TOFCs. During the entire duration of the flood, NUMMI did not have a minute of downtime related to parts coming from Chicago.

15.4.2. Consider dual-sourcing critical parts and products

Dual sourcing means using two suppliers for a component, raw material, product, or service. Note that the term "dual sourcing" is also often used when there are more than 2 suppliers. The idea is simple: If something happens to a supplier's site, the company can be served by the other supplier. This practice is widespread, often for the pure economical reason of playing the suppliers against each other in negotiations to drive down costs. The latter practice can be counterproductive in times of a crisis since it can result in low trust between suppliers and buyers.

> **Stop and Think 15.4**
> What are the advantages and disadvantages of busing the same product from multiple suppliers?

Don't jump to the false conclusion that the suppliers have to be two different legal companies. In fact, if the only reason for dual sourcing is risk hedging, a company only needs to ensure that a supplier can supply from different sites and that these sites do not use the same suppliers. This allows both a collaborative supply chain relationship, fewer relationships to manage, and lowered risk. It is the model of choice for Toyota.

The dual-source strategy in the meaning of having two different *legal* suppliers, however, makes supply chain and quality management more complex and alters the nature of the customer–supplier relationship. It reduces the amount of shared information and represents a lower level of mutual commitment. As we will see later in the example of Toyota's response to the Aisin Fire in 1997, it can be greatly beneficial to have a buyer–supplier relationship based on mutual trust and cooperation. It makes a difference, particularly in emergencies.

Dual-sourcing does not only apply to suppliers but also to the focal firm. If you produce all your products on one site, you are vulnerable if this site is closed for reasons like fire, flood, strike, contamination, war, transportation block, or trade sanctions. For the same reason that a focal firm would use dual strategies, your customer may want it too. If you can offer supply resilience, you not only increase the chance you may win parts of a sale opportunity but the whole contract.

An alternative to dual-sourcing is suggested by Takahiro Fujimoto,[19] an economist from Tokyo University. His "virtual dualization" or "virtual dual sourcing" strategy is discussed in Box 15.9. What it basically suggests is to not invest in capacity, multiple suppliers, or inventory, but instead build the capability to quickly set up or outsource needed capacity in times of crisis by having all design information ready for digital transfer. These ideas are worth considering, but they are not straightforward to implement and impossible to do for all a company's products.

19 Fujimoto, T. (1999). *Evolution of manufacturing systems at Toyota*. Oxford University Press.

Box 15.9 **Virtual dual sourcing**

In 2011, following the Fukushima earthquake, Takahiro Fujimoto formulated the concept of "*virtual dual sourcing*" of the supply chain as a strategy to pursue both competitiveness and robustness in the face of earthquakes, fires, floods, and other natural disasters. Fujimoto bases his recommendations on examples from the history of Toyota. His analysis is for a *local* disaster to which the global supply chain can respond. It did not envision a pandemic.

Virtual dual sourcing emphasizes the information infrastructure needed to produce a part, such as product and process designs, supplier lists, and technical details. When the crisis hits, the idea is to carefully and quickly choose between two options: Process restoration on the spot whenever there is time or the emergency transfer of critical design information to the backup production line until the damaged line is restored. Fujimoto rejects alternative strategies that involve holding inventory, spare capacity, or use multiple suppliers for the same parts.

In general, developing virtual duals poses problems of protection of intellectual property and technical feasibility. Within a supply chain, the design information may be owned entirely by the customer, entirely by the supplier, or parts of it by each. For example, the product design may be the customer's and the process design the supplier's. The customer is free to share what it owns with more than one supplier, but then it becomes multiple sourcing and each supplier's level of commitment is lower than with single-sourcing. A supplier may decide to share design information it owns with a competitor, but it is not risk-free, no matter how honest this competitor may be.

Once the companies work out the intellectual property issues, technical concerns remain. Starting and ramping up production in a line that hasn't previously been used for it is usually not a matter of days. Maintaining virtual duals ready to go for every item a supplier makes would be a drain on resources that, contrary to Fujimoto's stated goal, would weaken competitiveness.

While Fujimoto outlines a strategy, he does not give recipes that can be followed, as such generic recipes cannot exist for every manufacturer to respond to every possible emergency.

15.4.3. **Increase safety stock of critical parts**

Safety stocks are needed to hedge for long lead times and variations in supply and demand. Some see inventory as the answer to all supply chain problems, but it is not. Inventory usually creates more problems than it solves but not in the case of a shortage of critical parts. Technically speaking, all parts of a product can be equally "critical," but here criticality means "*time to recovery*" of part supply.[20]

20 Simchi-Levi, D., Schmidt, W., Wei, Y., Zhang, P. Y., Combs, K., Ge, Y., Gusekhin, O., Sanders, M., & Zhang, D. (2015). Identifying risks and mitigating disruptions in the automotive supply chain. *Interfaces, 45*(5), 375–390.

When emergencies affect suppliers' ability to produce and deliver, part supply can be recovered via switching to dual-source supplier, buying from new suppliers, finding alternative product designs, insource production of the parts, or waiting until the supplier can deliver again. All alternatives have different lead times. You want your safety stock to bridge the time between disruption and getting back to normal. This way, you can continue production during the suppliers' emergency.

For example, during the US–China trade war in 2019, the Chinese telecoms supplier and phone manufacturer Huawei Technologies decided to build a year of stock of crucial components in preparation for trade war uncertainties. Within one year – the planned time to recover – Huawei planned to develop alternative local suppliers. In this case, the decision may have come too late. Huawei was added to a list that made it illegal for any organization that operates in the United States to trade with Huawaei.

Another example is given in Box 15.10, which outlines how Toyota navigated the global chip shortage in 2020 by contracting suppliers to hold safety stock for Toyota.

Box 15.10 Toyota and the global chip shortage from 2020 onwards

In 2020, a global chip shortage made automobile factories all over the world come to a halt. A small chip effectively stopped the production of large cars and vehicles. One automobile company had, however, foreseen the shortage and stockpiled semiconductors for its products. Surprisingly for many, that company was Toyota – a company that had become famous for holding little inventory and still seeking to reduce it.

During the COVID-19 pandemic, the demand for digital products soared, semiconductor production slowed down due to infection risks, and global supply chains crumbled. The result was, among many other things, the "global chip shortage" forcing car makers to stop their lines or build semi-finished vehicles for rework in 2020.

Toyota, however, continued its production as normal into 2021. Toyota had learned from the Fukushima accidents in 2011 that some parts are worth stockpiling. Unlike many parts in a car, semiconductors are used in many industries, which considerably reduces the buying power of carmakers. Toyota asked its suppliers to stockpile 2–6 months' worth of chips for them.

Eventually, by August 2021, even Toyota started to run out of semiconductors and cut their global production capacity by 40%. Their production volume in September 2021 was 540,000 new vehicles compared to the planned 900,000. It affected 27 of 28 global production lines in 14 factories. In this case, the safety stock was not enough to bridge the "time to recover," but it helped Toyota outperform all its peers for a long period.

Stop and Think 15.5
Have there been events where you have hoarded critical supplies like toilet paper and food items? Why? Was it necessary?

15.4.4. Increase supply chain vigilance

Rather than inventory, a more effective strategy is to protect the supply chain by *vigilance*. This means keeping accurate inventory data, monitoring the in- and out-flows, monitoring the disruptions that can be anticipated, and responding quickly to events. Vigilance is the ability of a system to detect and interpret weak signals and alerts.

While the supply chain management literature says that a company should know its suppliers' suppliers' suppliers and its customers' customers' customers, the reality is that visibility rarely extends beyond Tier 1. Modern IT systems and software can increase supply chain transparency and hence provide more opportunity to catch signals. For example, those companies that used software systems to track what products they held in what locations during the onset of the COVID-19 pandemic, generally performed better than competitors with a more manual approach to supply chain management.

15.4.5. Increase supply chain resilience

Vigilance is the preceding step to *resilience*, which is the ability of a company to withstand disruption and minimize negative effects of an emergency. *Internally*, it is continuous improvement that develops the necessary resourcefulness to innovate in times of crisis. In *supply chains*, collaborative relationships between customers and suppliers make it possible to mobilize the entire network to recover from the disaster.

When a disaster hits your supply chain, the quick response cannot be yours alone. You need your suppliers' help, and that is why you cannot be in adversarial relationships with them. Long-term, single-source agreements, the regular exchange of business and technical information, and collaborative problem-solving are all necessary to cement the relationships that make a joint emergency response possible.

Box 15.10 shows how Toyota acted when they lost 99% of the supply of a critical component due to the Aisin Seiki fire in 1997.[21] Critics used the fire to point out Toyota's fragile supply system and declared failure of the TPS. However, rather than a failure of its system, Toyota considered this recovery as vindication of its approach to supply chain management, as the recovery was faster than anticipated in the press. Toyota's managers also realized that the recovery would have been easier if they had used 50 rather than 200 models of p-valves, a level of variety for which they could see no justification. They had fallen victim to part proliferation, a problem that has become more common in manufacturing since Computer-Aided Design (CAD) has made it easier for engineers to design new parts than to reuse existing ones.

Emergencies may also expose errors in supply chain design. The COVID-19 pandemic of 2020, for example, revealed that it is unwise to outsource surgical masks to the other end of the world to save on price. The pandemic turned them from a commodity to a strategic good needed in massive amounts everywhere at once. As it turns out, the combination of vigilance and resilience that make it possible to enlist the supply chain in rapid recovery works better than inventory. In the case of the Fukushima earthquake, more inventory would simply have meant more losses.

21 Nishiguchi, T. & Beaudet, A. (1998). The Toyota group and the Aisin fire. *Sloan Management Review, 40*(1), 49–60.

Box 15.11 **Toyota's response to the Aisin Seiki fire in 1997**

Aisin Seiki made 99% of the brake fluid proportioning valves (p-valves) used by Toyota in Japan. It was a close supplier–customer relationship. Toyota owned 23% of Aisin Seiki's equity, bought 80% of the company's output, and the CEO of Aisin Seiki was a Toyoda family member. Aisin Seiki's Kariya plant was the single source for 200 varieties of p-valves used in 19 factories, each of which held about 4 hours' worth of inventory on these parts. P-valves are safety-related parts in cars. Figure 15.21 shows a few models used at the time.



Figure 15.21 Examples of p-valves made by Aisin Seiki in 1997

On February 1, 1997, the Aisin Seiki plant in Kariya burned to the ground. Four hours later all Toyota assembly plants in Japan stopped, and domestic production went from 14,000 cars/day to 0. As this accounted for 6% of Japan's GDP, the news media paid attention. Initial reports painted the event as fatal to Toyota's production system. Within 5 days, however, production restarted. Within 6 weeks, it was back to 14,000 cars/day, and the total cost of the event to Toyota was on the order of $200M.

Toyota's first response was to dispatch 400 engineers to Aisin Seiki and set up a *damage control war room* at Aisin Seiki headquarters. Their initial conclusions were unacceptable. It would take 2 weeks to restore partial production on a few milling machines and 6 months to procure the new machines needed. The 200 p-valve variations required custom jigs and tools for tapers, and 30 kinds of cutting tools.

To reduce the recovery time, Toyota scrambled to collect all the blueprints and other master data about the p-valves and the tooling systems, and to ask suppliers to set up makeshift production lines. This launched a competition among 36 suppliers with 150 subcontractors to be the first to deliver p-valves and allow production to restart. Within 5 days of the fire, 50 new lines were making small batches of p-valves.

The winner was Somic Ishikawa, a supplier of brake parts and suspension joints. Four engineers immediately started on 40 hours of non-stop work on jig design and had the jigs made by Meiko Machinery, which assigned 30 workers to work 3 shifts/day on this task. Technical and administrative staff at Somic Ishikawa operated the machines and after only 4 days delivered its first p-valves to Toyota. Other suppliers also came to the rescue:

- Taiho Kogyo: A bearing maker.
- Brother Industries: A maker of sewing and FAX machines made 40 p-valves/day.
- Nisshin Kogyo: The maker of the 1% of Toyota p-valves not made by Aisin Seiki.

15.4.6. **Build fire-fighting capabilities but don't worship them**

Not all emergencies can be solved by outsourcing activities to external partners. Instead, they have to be dealt with there and then in "the heat of the moment" by dedicated teams that improvise in all possible ways to avoid escalation of a crisis and eventually bring the production system back to steady-state.

This is the type of problem-solving that humankind has always admired. Consider the 1995 *Apollo 13* movie. It is a story of emergency response during the Apollo 13 mission in 1970 where an explosion crippled the spacecraft. "Houston, we have a problem!" is a paraphrasing of the original quote of astronauts Jack Swigert and Jim Lovell that people still use when they encounter unforeseen problems that may jeopardize whatever they're working on. The drama we enjoy for entertainment, however, can be a dangerous addiction in the workplace. To an average visitor, a well-run factory *is* nothing spectacular and its people appear to be working at a leisurely pace. A poorly run factory, on the other hand, offers permanent drama, with managers rushing around from one emergency to another. If real firefighters worked like that, we can only imagine how it would hurt their performance.

This mode of operation is commonly called "fire-fighting." As the rewards for putting out fires are larger than for keeping them from happening,[22] a fire-fighting culture is difficult to change. The challenge is to make factory operations routine and continually improve them, while developing and maintaining the ability to improvise an emergency response as needed. For example, as shown in Box 15.12, after the Fukushima earthquake in 2011, Renesas needed fire-fighting capabilities to rebuild the factory and return to normal operations.

Box 15.12 **Renesas semiconductor plant and the Fukushima earthquake**

On March 11, 2011, the Fukushima earthquake devastated the Renesas semiconductor plant, 80 km South of Fukushima. While the news coverage was focused on the nuclear accident at Fukushima, the wreckage of the chip plant disrupted the supply chains not just of cars but also of computers, mobile phones, and appliances, all of which used Renesas flashable microcontrollers.

Since the semiconductor fabrication process require special equipment and special skills that suppliers of other parts could not possibly acquire in a few weeks, the only option was to repair and restart the plant. Renesas got help from Japan's government and some of its own customers, including Toyota. Kazumi Nishikawa, an official at Japan's Ministry of Economy, Trade, and Industry explained, "It is important to support the company's efforts to recover from the fire as it could have economic and social impacts in Japan."

The equipment was running by April 30, 2011, and full production restored by August 31, 2011 – less than half a year after the earthquake.

22 Repenning, N. P. & Sterman, J. D. (2001). Nobody ever gets credit for fixing problems that never happened: Creating and sustaining process improvement. *California Management Review, 43*(4), 64–88.

15.4.7. **Manufacturing repurposing**

The COVID-19 pandemic that hit globally in 2020 caused manufacturing activity to halt except in industries directly related to it, like pharmaceuticals, disinfectants, personal protective equipment (PPE), test kits, or ventilators, where it caused a surge in demand. The other examples in this chapter were all *local*, and the affected companies had to respond to the disruption in one or a few links in their supply network. COVID-19, on the other hand, was *global*, and required a global response. As it affects multiple competitors, the misfortune of one is not an opportunity for others.

In a pandemic, outbreaks occur in different areas at different times. When factories close in area A, it creates shortages for its customers in area B. Two months later, as factories reopen in area A, an outbreak is shutting down their customers in area B. This is what happened with COVID-19 in the first half of 2020 between China, the US, and Europe. With COVID-19, the world is in uncharted territory. The methods developed to overcome earlier, smaller disasters are useful starting points but not solutions.

The demand for some products skyrocket during global crises. During wartime it is weapons, ammunition, military vehicles, and oil. In a pandemic, it is protective equipment, hospital beds, and relevant healthcare technologies. Since the possibilities to ramp up existing capacity enough to meet demand is limited, one opportunity is to repurpose other manufacturing capacities. This is known as *manufacturing repurposing* (Box 15.13).

Figure 15.22 shows 3D printers imported from Germany and deployed inside a Paris hospital 24 hours after the order on March 31, 2020. The Stratasys 3D printers were used to produce plastic holders for face shields. 3D printers are not the fastest way to make plastic parts but the easiest to set up and to repurpose once the demand subsides. This makeshift factory helped serve a local peak demand.



Figure 15.22 3D printers making face shield frames at the Cochin hospital in Paris, France (Credit: Stratasys)

Box 15.13 **Repurposing factories during a crisis**

Factory repurposing is a rapid response solution to address shortages of critical items that can protect or save lives by using freed-up or added manufacturing capacity. It can be a business model for some, but during the early period of COVID-19 it was mostly driven by corporate philanthropy – the desire of people to help other people in times of crisis. It is a source of goodwill.

The COVID-19 pandemic in 2020 motivated many manufacturers to help produce needed medical equipment. Producers of alcoholic beverages started to produce hand sanitizer, producers of fabric started to sew face masks and hospital clothing, and automobile manufacturers started to produce medical ventilators. Companies such as Apple, Dyson, GE, Toyota, and Tesla were all repurposing parts of their factories for completely other products than they used to produce.

In Italy – a country that was hit particularly hard by COVID-19 – famous brands rushed to help. Gucci produced facemasks. Race car specialist Ferrari partnered with ventilator manufacturer Siare Engineering International to help them quadruple weekly output from 150 to 500 ventilators, turning their headquarter factory in Maranello into a ventilator assembly factory. Lamborghini changed its upholstery department to produce surgical masks and protective shields for a hospital in Bologna. Giorgio Armani turned all of its Italian factories into manufacturing sites for single-use hospital medical overalls. Famous luxury brand Prada – with 19 of its 22 factories in Italy – produced thousands of hospital gowns and surgical masks.

Repurposing of factories only happens during crises. It is meant to be a temporary strategy. During World War II, factories in all involved nations were repurposed – often by law – to produce military equipment. The most famous of these factories is arguably Ford's factory in Willow Run, which was in fact not repurposed, but custom-built by Ford to produce bombers. During its peak production in 1944, employees in 2 9-hour shifts, working 6 days a week, produced 453 B24 airplanes in 468 hours. Hence, it was famous for having a takt time of 63 minutes, or "a bomber an hour." It closed down in 1945. By then it had produced 8,685 bombers!

Manufacturing repurposing is often much more difficult than politicians imagine. Product designs must be developed, tested, and approved by various regulatory bodies. Production lines need to be built and calibrated. Quality control and assurance routines must be implemented. Staff need to be trained and skilled. Legal hurdles must be tackled. And stable supplies of material must be secured. Each of these are massive challenges on their own. Even for simple products such as face shields and face masks, manufacturing and supply chain coordination represent significant challenges even for world-class manufacturers.

Summary

The most essential points in this chapter were as follows:

1. Performance measurement and management is omnipresent in modern business.
2. Performance management must align with and support the strategy.
3. Measures are only proxy representations of reality, and must be dealt with accordingly.
4. Manufacturers need a mix of leading indicators, which point to future outcomes, and lagging indicators, which represent past performance.
5. On the manufacturing floor, the language of things is more important than the language of money.
6. Manufacturers should use a balanced set of metrics covering, depending on strategic relevance, all or some of the performance areas: Safety, quality, delivery, cost, productivity, environment, and people/morale.
7. Performance boards visually communicate performance-related metrics to the people who can influence them.
8. Daily meeting structures is an integrated part of lean transformations and help cascade deviations from plan quickly through the organization so they can be solved fast.
9. The trend is towards digital performance boards, but the economic benefits compared to analog boards is still unclear.
10. Managers must understand that humans have incentives to game metrics and seek to reduce metric gaming behaviors.
11. Emergencies are inevitable and must be managed. Tactics that help reduce negative effects of emergencies are as follows:
    - Scenario planning
    - Dual source strategy for critical parts and products
    - Safety stock of critical parts, but not of all parts
    - Supply chain vigilance
    - Supply chain resilience
    - Fire-fighting capabilities
    - Consider manufacturing repurposing

Test yourself

1. What are the differences and relations among measurements, metrics, and KPIs?
2. What is the difference between lead metrics and lag metrics?
3. What are some of the main challenges related to Management by Objectives?
4. Define two typical metrics for each of the following performance areas:

- Safety
- Quality
- Delivery
- Cost
- Productivity
- Environment/sustainability
- People/morale

5 What is the purpose of the Balanced Scorecard?

6 What is Activity-Based Costing and what problem is it trying to solve?

7 Quickly sketch a possible layout for a typical shop floor performance board and suggest reports to be included.

8 List the main pros and cons of digital shop floor boards versus analog boards.

9 Describe how daily management meetings cascade through a lean organization.

10 Discuss at least five ways to mitigate metric gaming.

11 Suggest three tactics to protect a company against supply chain shortages during crises.

12 What is manufacturing repurposing?

Further reading

Armstrong, M. (2021). *Performance Management: Key strategies and practical guidelines* (3rd ed.). Kogan Page.

Chernov, D. & Sornette, D. (2016). *Man-made catastrophes and risk information concealment*. Springer.

Drucker, P. (2012). *Management*. Taylor & Francis.

Ferdows, K. & De Meyer, A. (1990). Lasting improvements in manufacturing performance: In search of a new theory. *Journal of Operations Management, Vol. 9*(2), 168–184.

Gurnani, H., Mehrotra, A., & Ray, S. (Eds.). (2012). *Supply chain disruptions: Theory and practice of managing risk*. Springer.

Manners-Bell, J. (2020). Supply chain risk management: How to design and manage resilient supply chains. Kogan Page Publishers.

Merchant, K. A. & van der Stede, W. A. (2012). *Management control systems: Performance measurement, evaluation and incentives*. Prentice Hall.

Monden, Yasuhiro (Ed.). (2013). Management of enterprise crises in Japan. *Japanese Management and International Studies*, 10.

Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design: A literature review and research agenda. *International Journal of Operations & Production Management, 15*(4), 80–116.

Neely, A. (2005). The evolution of performance measurement research. Developments in the last decade and a research agenda for the next. *International Journal of Operations & Production Management, 25*(12), 1264–1277.

Norrman, A. & Wieland, A. (2020). The development of supply chain risk management over time: Revisiting Ericsson. *International Journal of Physical Distribution & Logistics Management, 50*(6), 641–666.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

16 Improving Performance

According to an old maxim, "the biggest room in the world is the room for improvement." That is certainly true in factories. The ability to continually improve processes is a hallmark of successful manufacturers. Improving performance is an objective always worth pursuing, but it is far from easy. It is not unusual to improve performance in one dimension while degrading it in other dimensions. Instead of making things better, "improvements" can make things worse – known as *kaiaku* (改悪) in Japanese and, jokingly, Verschlimmbesserung in German. Besides wasting resources before, during, and after making them, they are detrimental to the development of an improvement culture.

The improvement of performance, however, can be managed. Improvements are often the result of dedicated projects seeking to solve an identified problem. There are proven methods and templates that help manufacturers identify and improve performance. Toyota calls all improvements *kaizen* (改善) regardless of scale. Others in Japan reserve kaizen for small and incremental improvement projects, usually in large numbers, and use kaikaku (改革) for larger and more radical improvement. There are many helpful templates for small and incremental improvements. Larger projects, on the other hand, require a case-by-case analysis starting from goals and available resources.

To coordinate across multiple sites, large corporations often use performance improvement *programs*. While projects and programs are helpful structuring mechanisms, the much greater challenge is to develop a *culture* of continuous improvement as part of everyday work. When all improvements are delegated to a special department, the pace of improvement is only a fraction of what it can be when everyone participates. This chapter is an introduction to performance improvement. It covers the essential concepts and how to organize and manage improvement projects and programs. A range of useful methods and tools are covered in Chapter 17.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Recognize improvement potentials in manufacturing.
2. Differentiate between kaizen and kaizen events.
3. Be familiar with the most common project-management methods.
4. Set up and use an obeya.
5. Discuss the characteristics of key production improvement models.
6. Understand the main ideas of the Toyota Production System and lean production.
7. Develop a company-specific production system.
8. Recognize the potential of new technologies to improve performance.
9. Understand the role of leadership in the development of an improvement culture.

DOI:10.4324/9781351110310-21

16.1. The case for improvement

A factory can always be improved. Once you have achieved any level of performance, it is just the starting point for the next round of improvement. Perfection is something you never achieve but always pursue and, if you dig deep enough, you always find opportunities. While "optimization" is a frequently used word in manufacturing, managers should be aware that an optimum only exists in a mathematical model.

When we say that a factory can always be improved it is a *postulate* you make when you walk through the gates. The overwhelming empirical evidence is that, if you assume it, you find opportunities.

Some believe that manufacturing improvement is futile and a distraction from more "strategic" pursuits like outsourcing or product innovation. It is a usual misconception that factories are "close to optimal" and any additional improvement will yield diminishing returns. These people believe that, as you keep reaching for fruits that hang ever higher, the effort becomes excessive for the benefits. However, there are at least four things to consider:

1. **Competition is constantly raising the bar.** If your competitors keep improving and you don't, you lose.
2. **The market is constantly developing.** Customers may ask for price reduction, shorter delivery times, or custom products.
3. **Technology is constantly evolving.** New technologies offer new opportunities to improve, incrementally or radically. Practitioners of continuous improvement are better informed, savvier buyers of new technology.
4. **Improvement means learning.** As you make improvements, you enhance not only performance but also your own skills, so that some of the potential that was out of reach before no longer is. The fruits may be hanging higher, but you are a better climber.
5. **The environment is changing.** Pandemics, wars, climate change, geopolitical changes, financial crises, and other unforeseen events challenge your ability to adapt.

For these reasons, if you do not improve your manufacturing operations, you will lose ground to competitors, providers of substitute products, or new entrants, as shown in Figure 16.1.

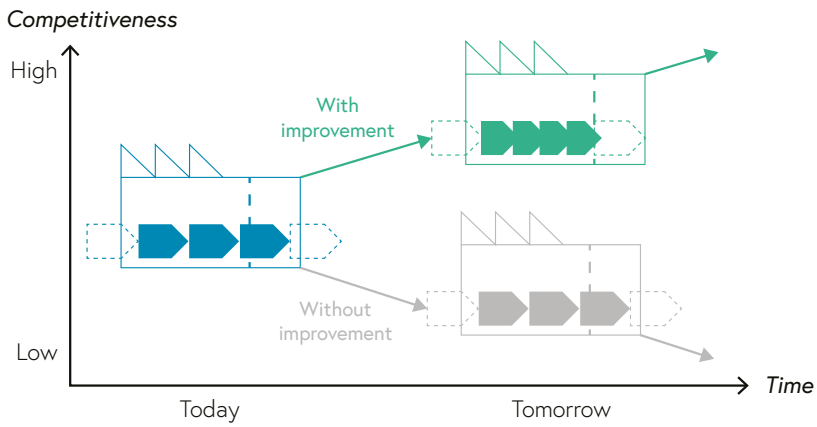


Figure 16.1 Improvement is the only way to sustained competitiveness

The best manufacturers are able to both continuously improve their existing processes and create new processes and products. This rare ability is called *organizational ambidexterity*,[1] and it means that they find a balance between exploitation of current ways of working and exploration of new ways.

Some argue that companies pursuing continuous improvement will be slower to develop and adopt disruptive technology. But does a machinist who improves a fixture deter the founder of the next startup? There is no connection. If the machinist were not making improvements, his or her creativity would be untapped. The improvement work does not divert venture capital from disruptive technology. This is an example of *structural ambidexterity*, where radical innovation is done by another group than continuous improvement.

In reality, companies that have created an improvement and performance culture are likely to be more innovative overall, with more employees bringing forward both incremental and radical suggestions. The challenge is to create and maintain organizational routines for capturing, evaluating, and implementing them.

16.2. **Kaizen**

Kaizen (改善) is a Japanese term that means "change for the better," and is often translated to "continuous improvement." Since, in this context, "continuous" means "done all the time," "continual" might be more descriptive than "continuous." However, we stick to the more established "continuous improvement."

Continuous improvement most commonly refers to small changes to work methods, conceived and implemented by the people who do the work. Every employee, regardless of rank, is expected to spend part of the work time on improvement, and, when moving onto another assignment, should leave the workplace, department, or division performing better than he or she found it. Whatever you are in charge of, you are not just responsible for running it, but for making it better. Hence, employees have two jobs:

- Do the work
- Improve the way it's done

Besides its meaning within TPS, even inside Japan, kaizen is interpreted differently by different authors like Masaaki Imai[2] or Bunji Tozawa and Norman Bodek.[3] Tozawa and Bodek define kaizen as "changes in methods to make the work easier, conceived and implemented by those who do it." To Imai it is "ongoing improvement involving everybody, without spending much money." To Tozawa and Bodek, each discrete change is a kaizen; to Imai, kaizen is the process by which these improvements are made on a continuing basis. As defined in the *Kojien* (広辞苑), the Japanese equivalent of Webster's dictionary, kaizen (改善) is "the act of making a bad place good again."

1 Tushman, M. L. & C. A. O'Reilly (1996). The ambidextrous organization: Managing evolutionary and revolutionary change. *California Management Review, 38*/(4), 1–23. / Benner, M. J. & Tushman., M. L. (2003). Exploitation, exploration, and process management: The productivity dilemma revisited. *Academy of Management Review, 28*(2), 238–256.

2 Imai, M. (1986). *Kaizen (Ky'zen), the key to Japan's competitive success.* Random House.

3 Tozawa, B. & Bodek, N. (2001). *The idea generator: Quick and easy kaizen.* PCS Press.

16.2.1. **Kaizen events versus kaizen**

In the lean literature, kaizen is often misrepresented as "*kaizen events*," an improvement project format developed in the US. Kaizen events are not really performing continuous improvement. As an oxymoron, "*kaizen blitz*" is even worse: It mixes Japanese and German in a concoction that literally means "lightning strike of continuous improvement." Terms like "kaizen event" or "kaizen blitz" mislead because they use the term kaizen but what these activities do is not kaizen. The main problem is that, by monopolizing attention, these terms make it nearly impossible to promote, organize, or even discuss actual kaizen activity.

> **Definition**:
> A kaizen blitz is a rapid improvement workshop designed to solve discrete process problems within a few days.

According to an eponymous book, in the early 1990s, members of the Association for Manufacturing Excellence (AME) were frustrated by lack of progress in lean implementation using Quality Circles, and developed the *kaizen blitz* as an alternative.[4] The visible part of a kaizen event is a period of 5 days during which a team designs and implements changes in a targeted area of the factory. The team includes members from the production team of that area and from support departments. The team is dedicated for the 5 days but calls upon others as needed, to move heavy machinery, or to validate the safety of the design.

The concentrated 5-day effort provides the appearance of rapid results. As a project, however, a kaizen event takes about 3 months. Planning, scheduling, and preparation takes about 6–8 weeks upfront, and the 5-day event usually leaves a list of action items needed for finishing that requires another 6–8 weeks.

Many manufacturers have adopted the kaizen event as a panacea, ignoring all other methods. As a result, improvements that don't fit within the kaizen event format didn't happen. In companies that set quantitative targets such as "the number of kaizen events per month," kaizen events quickly degenerate into formal exercises that accomplish little. The kaizen event works when applied to the right opportunities, but there are two problems with it:

1. **The kaizen event's promise of instant gratification has made it so popular that all other means of changing work are forgotten.** It is a problem because it leads organizations to ignore opportunities that are either too small or too large.
2. **The kaizen event takes attention away from real kaizen.** There is abundant Japanese literature that makes no reference to anything resembling kaizen events. In fact, the improvements that are called kaizen are too small for kaizen events.

Kaizen events have a similarity to what is known within Toyota as *jishuken* (自主研), literally "autonomous study." Jishuken is a richer and more varied approach than the kaizen event. It started as just study groups on TPS about 1976, primarily for the purpose of training suppliers, along the lines of the self-study seminars that the Japanese Union of Scientists and Engineers (JUSE) had been organizing for decades. For example, a jishuken group of suppliers would involve 2–3 members from each of 6–7 non-competing suppliers, conducting improvement projects in rotation at each of the participating companies. Each project lasts 8 weeks, involving multiple visits by the outside members and homework between visits by the inhouse members. Jishuken is a richer approach than the *kaizen blitz*.

4 Laraia, A. C., Moody, P. E., & Hall, R. W. (1999). *The kaizen blitz: Accelerating breakthroughs in productivity and performance*. Wiley.

16.2.2. The value of small improvements

Many hotels are good places to experience the lack of kaizen. A hotel that does not practice kaizen may have a slippery marble lobby. If you enter on a rainy day, your wet shoe may slip and you may bang your knee on the marble. Marble looks great. That it is slippery when wet is a detail that had not been considered at design time. The same lobby can have a luxurious restroom with a door that does not automatically unlock when you leave, so that you could leave it both vacant and locked. In your room, there may be 1 light bulb not working, or only 1 toilet paper roll with too few sheets for a full day. The big-picture design of the hotel can be fitting for a prestigious company, but the details are botched.

There are, however, too many details for managers, engineers, or architects to worry about. This is why companies seeking excellence need kaizen. They need the help of the people who are affected by these details to fix them. When, in a company, you notice small things done in clever ways, it is a telltale sign of kaizen activity. It enhances skills, builds confidence, and improves morale, but the key reason you do it is to make sure sloppy details don't hurt your business. This is what kaizen means in Japan.



Figure 16.2 Kaizen example

Figure 16.2 shows an example of genuine kaizen in a US factory. The workpiece in the vise is 28 feet long, and requires greasing in multiple locations. The operator on the left was tired of running back and forth to a fixed location to pick up the grease. The cart now contains everything he needs to apply grease anywhere on the work piece, and he wheels it back and forth as needed. To the right is the production supervisor for the area, who supports this and other similar projects.

How was it actually done? The production team from this area was given a budget of $500/operator to spend as they saw fit on supplies and devices for improvement projects at a Home Depot store. Their actual spend worked out to $113/operator, including the cart and bins you see on the picture and a magnetic sweeper. It is a perfect illustration of the kaizen concept. It is too small an improvement to warrant the attention of engineers or managers, yet it makes the work easier for the operator and makes him more productive.

The only way to make sure such improvements are made is to enable and encourage the people who do the work so that they do it themselves. It is a valuable part of TPS, but it is not all of it. Higher-level issues must also be addressed, include make-versus-buy decisions and production line layout.

> **Stop and Think 16.1**
> How do you think the operator in Figure 16.2 feels about this improvement? Does it matter whether the suggestion came from him or his supervisor?

Companies that are serious about kaizen often offer the infrastructure to engage in it routinely. An example of such infrastructure is the *kaizen shop*, a place where operators can try their own ideas without putting production at risk (see Box 16.1). This is sometimes also called a *sandbox*.

Box 16.1 **Kaizen shops**

A kaizen shop is a physical place where improvement ideas can be realized quickly and *by the person who had the idea*. Many companies deprive operators from the opportunity to realize their own ideas because they think their labor time is better spent on the assembly line. They try to source ideas from operators based on their first-hand experience of problems, then have Engineering build a customized solution that is provided back to the operator. This procedure is a waste of motivation and learning on the side of the operator. A better way is to help the operator realizes his or her own idea.

Figure 16.3 shows and example of a kaizen shop from Volvo Group's truck factory in New River Valley, USA. In this factory with 3,000 employees, there are 2 large kaizen shops where all operators can come and try their improvement suggestions. Standard and reusable building material exists to build boards, tables, carriers, or other equipment. A kaizen shop supervisor is available to help build and advise.



Figure 16.3 Kaizen shop in Volvo Trucks, New River Valley, USA

The kaizen shop is a vital space for helping organizations implement employee suggestions. Employees make suggestions during daily performance meetings or via physical or electronic suggestion boxes. Engineers and managers then consider either approving or returning these suggestions.

If a suggestion is returned, the author receives feedback on why, and encouragements to keep participating. If it is approved, its author can ask to be replaced for 4 hours at the normal workstation to implement the idea using the materials and help from the kaizen shop. The point is that, if you cannot implement suggestions, there is no need to collect them.

16.2.3. **Factory life with and without kaizen**

In *Kaizen*, Masaaki Imai describes Japanese executives returning in the 1970s to American plants they had visited 30 years before and being struck by the *absence* of change: They saw the same production lines with the same equipment operated the same way. Figure 16.4 shows the same work done the same way 30 years apart at Ford's River Rouge plant.



Figure 16.4 No visual change in River Rouge coke ovens 1946–1976

If there is kaizen activity in a factory, how does it change the work life of employees at all levels? Figure 16.5 compares the breakdowns of their use of time, with and without kaizen.

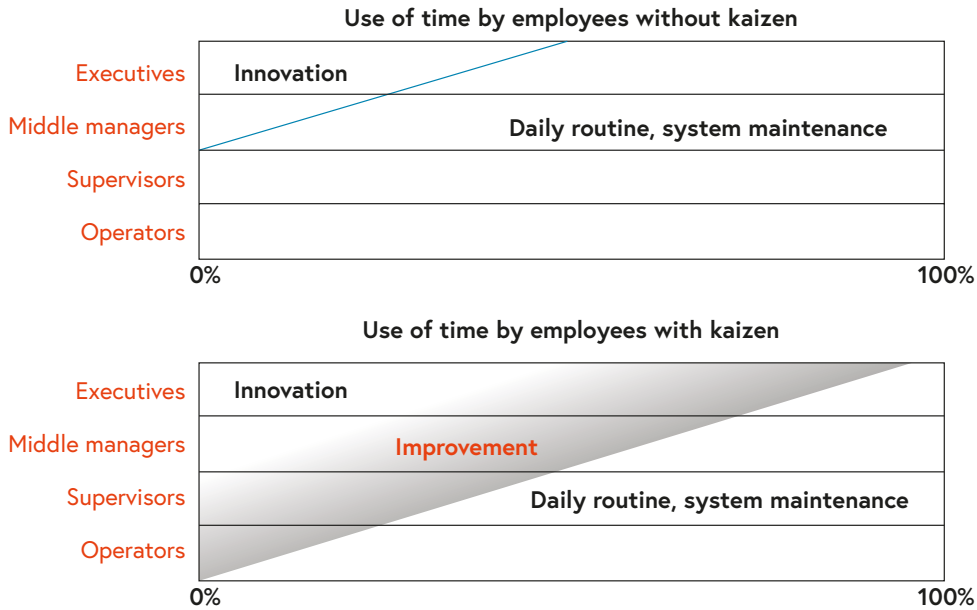


Figure 16.5 Time use with and without kaizen

In a plant without kaizen, operators and supervisors are fully occupied with routine daily tasks: Production for the operators, expediting parts, enforcing discipline, and record keeping for supervisors. Only middle managers and executives spend a fraction of their time on projects involving capacity changes,

new production lines, or new technology. Nobody works on incremental improvements to the way the work is done today.

By contrast, kaizen involves employees at all levels in such improvement activities to different degrees. Between improvement and daily routine, the boundary is sharp; between improvement and innovation, fuzzy. Over time, the cumulative effect of incremental improvement is radical change by itself. In addition, the skills acquired and the lessons learned from incremental improvements feed new line or plant designs.

In some Japanese auto parts plants, you can see automatic lines side-by-side, where one used old machines gradually retrofitted with devices that reduced the need for human intervention, while the other one had been built from scratch with new machines to be automatic. The kaizen work done on the former is essential to the success of the latter.

Kaizen applies to all jobs. An example of kaizen in cleaning jobs is given in Box 16.2.

Box 16.2 **Kaizen in cleaning jobs**

The nature of kaizen as practiced is best shown through examples. About 20 years ago, the monthly Japanese magazine *Kojo Kanri* (*Factory Management*) focused an issue on kaizen for the kind of jobs that do not receive much management attention. There was in particular a story about a circle of high-speed train janitors who were tired of cleaning the same toilets 8 times a day. These trains were equipped with traditional, Japanese squat toilets that international passengers did not know how to use and soiled as a result.

One obvious solution was to replace these toilets with the worldwide standard. It would have been no hardship for the Japanese passengers, as this style is already used in 90% of their homes and work places. But replacing these toilets in 100 16-carriage trains could not have been done quickly, and was an investment beyond the scope of kaizen.

So the janitors took a number of simpler steps, and measured the results in terms of the number of required cleanings per toilet per day. This included posting some graphic explanations on how to use the equipment, painting outlines of where users should place their feet, and finally materializing the right locations with rubber pads to make it awkward to place your feet anywhere else. At the end of the project, the required cleanings were down from 8 to 1 per toilet per day.

16.2.4. **Role of standards**

In factories, you frequently hear comments like "Our teams have good results with kaizen, but nothing seems to stick for long." This begs the question of what "good results that don't stick" are. Successfully completing a kaizen project means *reducing a new work method to daily practice*. A compelling demo in a management presentation is not the end. Claiming good results on this basis is like declaring victory after running a quarter of the race. By definition, if it doesn't stick, it's not a good result. If improvements erode systematically, then the whole kaizen activity is a failure.

How can we know if we have actually improved a process? Besides the requirement to stick, it also has to be better than the previous process. The only way to be sure about the latter is to measure before and after performance along the relevant performance dimensions, commonly: Safety, quality, delivery, cost, environment, and people (SQCDEP).

You can only measure a consistent process, which, in the lean literature, is referred to as a *work standard*. Effective standards are necessary to ensure that the same work is done the same way every time, by the same or by multiple operators. Without consistency in work processes, the measurement one day with one operator can be very different from the measurements another day with another operator. Large variations in measurements diminish our ability to clearly say if a change really is an improvement.

An often repeated idea is that a problem is a deviation from a standard, and you can't identify a problem unless you have standards. If you cannot identify a problem, you have no opportunity to improve. This idea is summarized by TPS pioneer Taiichi Ohno, who has been quoted as follows:

> *"Where there is no standard, there can be no kaizen."*

Often, however, the problem is the *absence* of any work standard, or of current standards that anyone adheres to. The key points of an operation are often only in the head of one experienced operator, and 5-year-old binders of obsolete instructions are no help. In such cases, with the permission of the operators, you can document current practices through video recordings that you use as a basis for improving them. Then you generate work standards.

Work standards are often depicted as an improvement backstop, a formalization of the improved process for the purpose of preventing backsliding. This is commonly illustrated as in Figure 16.6. The PDCA or PDSA cycle is discussed in Section 17.5.1 on page 722. As the cycle moves 1 full circle up, the team updates the job instructions and make it serve as a stopper to prevent backsliding.

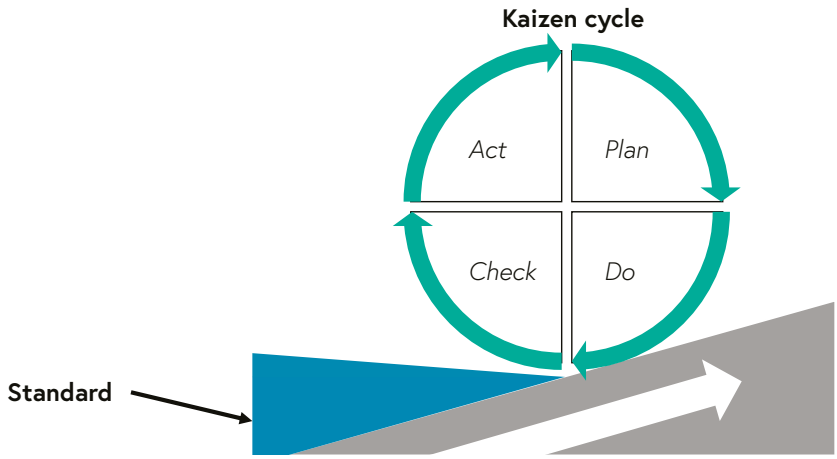


Figure 16.6 A common representation of the role of standards in kaizen

Figure 16.6, however, is simplistic because the analysis of a task often reveals that operators don't follow the work standards, either by mistake or because they have already found a better way. Even though the work standard is supposed to be the "best known way," it's not always the real starting point of improvement, contrary to the way Figure 16.6 pictures it.

Work standards are *documents*, and part of Master Data. Unless there is a 3×5 card, an A3 sheet with infographic, a one-point-lesson, a binder, a video, or any other kind of document specifying how something should be done, there is no standard. Undocumented habits or expectations are not standards. They are neither verifiable nor enforceable. Job instruction cards, as developed in the 1940s as part of the US government's Training Within Industry (TWI) program have been a particularly effective format (Box 16.3).[5] The state-of-the-art in 2022 is to create video instructions with the help of AI, as for example Stanley Black & Decker does with DeepHow.[6]

Box 16.3 **Training Within Industry (TWI)**

C. R. Dooley, W. Dietz, M. J. Kane, and W. Conover developed *Training Within Industry* (TWI) as the US War Department started in 1940 to rapidly train inexperienced recruits in factory work, crediting many ideas to Charles R. Allen. This was in anticipation of the mobilization of current male workers into World War II. The TWI program focused efforts on training and coaching supervisors.

The US government stopped supporting TWI in 1945. It lived on through the TWI Foundation, which updated TWI manuals and held conferences until 1952. The Chicago School of Business taught parts of TWI until 1970. TWI came to Japan by way of India, where the Japanese Ministry of Labor sent two people for British TWI training, organized through the International Labor Organization (ILO) in 1947. Returning to Japan with TWI manuals, they ran pilot projects at Nissan, the railroad, and a textile company. Among others, Toyota adopted it and brought it back to the US. 35 years later in its NUMMI joint venture with GM, reviving interest in its use.

TWI has 3 parts:

1. *Job instruction (JI).* JI is the on-the-job training of employees to execute the process. The trainer can be a first-line manager, a team leader, or a co-worker. The trainee is certified for each task when able to execute the sequence of steps while narrating the execution and the key points of each step.
2. *Job methods (JM).* JM is training for first-line managers, team leaders, and operators on process improvement.
3. *Job relations (JR).* JR trains first-line managers and team leaders to build trust, collaboration and engagement within their teams.

Each of the 3 J-programs entailed 5 consecutive days of training at 2 hours/day, for up to 8 supervisor per class. Instructor training was followed by coaching on using the new knowledge on real-life problems. The concepts of TWI have aged well. Indeed, the original documents from the 1940s could have been written in the 2020s. The pocket card carried around by Toyota supervisors is similar to the original pocket card distributed by the U.S. government in the 1940s. A question, however, is if one really needs a pocket card in the time of the smart phone. The training technology of the 1940s was limited to typescript on 3x5-in laminated cards; today, we can use smartphones and tablets to display instructions with pictures, sound, and video – or use Virtual Reality (VR) and Augmented Reality (AR).

5 Warren, M. (2010). *The TWI Report*. Lulu Press, Inc.

6 Fretty, P. (2020, December 10). Stanley Black & Decker realizes AI is a powerful tool. *Industry Week*.

16.2.5. Improvement tools, methods, and methodologies

Tools are devices we make or buy that enable us to do things we couldn't with our bare hands. To become adept at using a tool, we need to learn or create *methods*. When tools and methods are arranged in a fixed sequence to obtain a consistent outcome, the result is actionable instructions. Without instructions, we could still figure out how to do the task, but it would take longer and involve more errors.

There is more to playing the piano than practicing scales, but you can't get there unless you do practice scales. Likewise, there is more to improvement than tools but you need the tools. They are not *sufficient*, but they are *necessary*. Unless you learn them, you are depriving yourself of the benefits of clever tricks, methods, and analytical tools developed by others over decades. Each one doesn't necessarily take long to learn but would to reinvent, and you don't have that time.

Tools are things we buy or build, whose possession does not make us adept at using them. Methods are things we learn. The difference is clear in daily life but blurred in management-speak, where, for example, a method like SWOT analysis is often called a *tool*, even though all it has that resembles a tool is a 2×2 matrix on a white board with squares labeled "strengths," "weaknesses," "opportunities," and "threats." The value of a SWOT analysis depends exclusively on the skill of filling out this matrix in a brainstorming session. Also, in manufacturing, we often conflate tools and methods as, for example, in Kaoru Ishikawa's[7] "7 tools of quality control" (Box 16.4), which range from check sheets to control charts.

We all have a set of tools and methods that we choose to use when the need arises. This choice is strategic to problem-solving and requires thought. You have more than just a hammer, not every problem is a nail, and you may have more than one tool you can use. A *methodology* is quite different, a "12-step, proven process" that you surrender to and relieves you from the need to *think*, and from *responsibility*.

Methodologies are successful for many endeavors, such as baking a cake, building a house, or assembling a product. However, the improvement of a factory is often more like exploring uncharted territory than building a house, as in Figure 16.7.

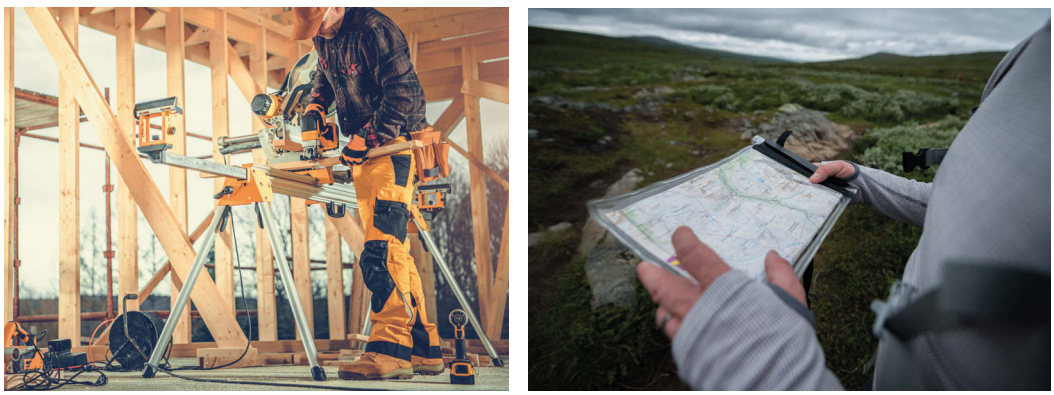


Figure 16.7 Two types of projects: Known situation (building a house) vs. unchartered territory (into the wild)

7 Ishikawa, K. (1968). *Guide to quality control*. Asian Productivity Organization, Tokyo.

When exploring uncharted territory, you don't know what lies beyond the next hill, and must constantly adapt what you are doing to what you currently know. Evidently, tools that are fit-for-task are useful in both situations – the saw and the map and compass –, but methodologies are more effective in known situations such as building a house. When navigating unchartered territory, one sometimes need to come up with new tools and methods. Just as with tools, there is no methodology that works in all situations.

There is no consensus on a toolbox to improve factories. When you google, for example "tools of lean," you get lists from 5 to 100 items from many sources, *some* of which do not pertain to manufacturing. John Bicheno and Matthias Holweg[8] and Radu Demestrescoux[9] have compiled toolbox books for lean production. Many of these "tools of lean" would be better described as methods that have emerged as point solutions to problems Toyota encountered over decades, and coalesced into the Toyota Production System, as described by Takahiro Fujimoto.[10]

Box 16.4 **Ishikawas' 7 tools of quality control**

Kaoru Ishikawa knew that statistics was key to improve manufacturing quality, but he also realized that most people struggled with it. One of his objectives was therefore to make statistical analysis less complicated for the average person. As a result, Ishikawa suggested the 7 basic tools of quality control, which, according to him, could solve "95% of quality related problems in the factory" – a number that was probably not based on data.[11]

The original 7 tools of quality control were as follows:

1. *Cause-and-effect diagram*: A diagram that looks like a fishbone where the problem being investigated is shown at the head and any potential underlying causes are investigated as the bones and sub-bones, usually separated into the four M categories human, method, materials, and machine (also known as Ishikawa diagram).
2. *Check sheet*: A data recording form where the operator makes a check for each type of error or event recorded.
3. *Control chart*: Shows the evolution over time of one quality characteristic measured after the operation is complete and issues alarms when it shows evidence of a change with an assignable cause.
4. *Histogram*: Shows the distribution of a sample of a numeric variable among bins.
5. *Pareto chart*: To tell the vital few from the trivial many among categories, for the purpose of identifying the most promising projects (also known as the 20-80 rule).
6. *Stratification*: Slicing and dicing the data based on a variety of criteria.
7. *Scatterplots*: Visualize relationships between two parameters in a chart.

The 7 tools of quality control are still widely taught but they are outdated in terms of effectiveness. The three first ones are the least trivial and least known and therefore covered in the next chapter.

8 Bicheno, J. & Holweg, M. (2009). *The lean toolbox*. PICSIE Books.

9 Demetrescoux, R. (2019). *La boîte à outils du lean* (2nd ed.). Dunod.

10 Fujimoto, T. (1999). *The evolution of a manufacturing system at Toyota*. Oxford University Press.

11 Antony, J., McDermott, O., & Sony, M. (2021). Revisiting Ishikawa's original seven basic tools of quality control: A global study and some new insights. *IEEE Transactions on Engineering Management*. In press.

16.2.6. **Wisdom**

The Toyota Way 2001[12] emphasized wisdom (chie, 知恵) but it is rarely discussed in the English-language literature on TPS or lean. Wisdom can be viewed as the ultimate step in the mastery of a process. In succession, you have the following:

- **Skill.** You can execute a task.
- **Know-how.** You can combine and sequence tasks to get a job done.
- **Know-why.** You understand the logic behind the tools and methods.
- **Wisdom.** You can adapt the tools as needed in new circumstances.

This notion of wisdom is the *resourcefulness due to the assimilation of accumulated knowledge*. It differs from *common sense*, which is resourcefulness too, but in a field where you have no specialized knowledge. If you have to study and experiment, then it's more than common sense. Common sense is what you fall back on when you must improvise a response to an emergency of a type you are not familiar with. An example of a specific improvement technique that requires wisdom is karakuri kaizen (Box 16.5).

Box 16.5 **Karakuri kaizen**

Karakuri kaizen refers to the use of materials handling devices that rely on gravity, levers, cams, and inertia to move bins in elaborate ways, transfer parts between machines, or deliver a controlled number of small parts to an operator's hand. Such devices have long been used as parts at Toyota and as part of lean thinking, but now we have a generic name for them. Figure 16.8 shows an example of a karakuri equipment built in Volvo AB in 2022.



Figure 16.8 Karakuri equipment for assembly work at Volvo AB (Credit: Serge Caron, Volvo AB)

The principles of karakuri kaizen are as follows:

1. Don't use the human hand to move objects
2. Don't spend money
3. Use the power of your equipment or nature
4. Build it with the wisdom and creativity of the people of the shop floor
5. For safety, build a device that stops automatically

12 Toyota Motor Corporation (2001). *The Toyota Way 2001*. Internal document, Toyota.

16.3. Improvement projects

Improvement activities vary in scope and requirements in employee engagement, and the methods used to plan and manage each project should be based on its needs. Projects may involve one individual, a local team, a task force drawn from multiple groups, or the entire organization. They may take one afternoon or 3 years, and managers must organize and oversee them accordingly. Factories should not make the mistake of limiting themselves to only one or a few ways of improving.

Over the years, templates have emerged for a variety of projects from small to large:

- **The "just-do-it."** It is informally managed between an operator and a team leader.
- **Suggestion systems.** Suggestion systems involve proposals submitted on forms by individuals, via boards, mail boxes, or electronic systems that are formally reviewed by management. They are often tied to reward and recognition schemes.
- **Quality Circles.** These are small groups of volunteers meeting weekly to work on a project taking 3 to 4 months.
- **Kaizen events. Also known as "kaizen blitz," they are centered on a week of intense, full-time activity by a team.** Each event requires about 6 weeks of preparation and another 6 weeks of follow-up after the week, adding up to about 3 months.
- **Task forces.** Larger projects, like new line or new plant designs, require a task force – that is, a team drawn from multiple departments for the duration of the projects. A task force may start small and grow in size as the project advances.
- **Long-term projects.** Projects involving sustained activity over months or years, like standardizing dimensions on 300 dies. These projects require fewer participants for longer periods, patiently whittling down the problem.
- **Plant-wide and company-wide activities.** These pose the special challenge of requiring participation by all members of the organization, including temporary workers. "Total" in translations of Japanese acronyms means "with participation by everyone."
- **Multi-company projects.** Joint projects with suppliers or customers require trust between different companies.

16.3.1. Suggestion systems are popular but rarely effective

Many improvements never happen unless operators make them happen but soliciting their help is not enough. Historically, the first attempt was the *suggestion system*, dating back to 1880. It is still in use at many companies, including Toyota, but, while it is part of continuous improvement, it is not an approach to systematic problem-solving.

Shop floor employees make suggestions about whatever they have ideas about, which can be a machine malfunction, ergonomics, the location of materials, safety, visibility, hygiene, break areas, work clothing, or others. All such thoughts are worth surfacing and considering but not all of them will lead to improvement of factory performance.

Unfortunately many suggestion systems don't deliver. Many manufacturers try to kick-start a lean program or continuous improvement effort with a suggestion system tied to tangible rewards, which can be anything from an iPad, two extra holidays, or a percentage of the savings from the improvement. The result can be an overflow of suggestions ranging from free coffee to replacing production machinery.

The system burdens management with the responsibility to (1) encourage participation, (2) review all suggestions, (3) respond to their originators, (4) recognize participation, and (5) implement the accepted suggestions. The result is often a dysfunctional suggestion box system as illustrated with the paper shredder in Figure 16.9.



Figure 16.9 Suggestion boxes usually don't work

Usually, suggestion systems break down for any of the following reasons:

- It is easy to start but hard to sustain.
- Too many suggestions not evaluated or rejected demotivate participants.
- Poor quality of suggestions.
- Long backlog of suggestions.
- Lack of implementation resources.
- Lack of skill in assessing the potential of every suggestion.
- Lack of employee ownership of improvement suggestions.
- Difficulty in assessing the overall effectiveness of the system.
- Gaming when rewards are substantial.
- Lack of ownership of the system.

Rewards for contributing suggestions and for having them implemented are necessary to *convey a message* of appreciation and encouragement. Recognition at town-hall meetings, plaques, articles in the newsletter, or in the local press are all *symbolic* recognition that must be consistent with the local culture. Some operators, for example, may be wary of being singled out by management in front of their peers, while others may welcome it.

Monetary awards are tempting to use because they create tangible incentives, but they are rarely sustainable and tend to create gaming behaviors. The best option is to never introduce them in relation to kaizen activities. They signal that providing improvement suggestions is *not* part of the regular job.[13]

Consistent improvement contributions must however be recognized in terms of salary increases, bonuses, or career opportunities. The important point is appreciation. In factories that do not have a continuous improvement culture, the suggestion system may be the only way to engage employees in improvement. It isn't optimal, but better than nothing and can be a start.

13 Netland, T., Schloetzer, J. D., & Ferdows, K., (2015). Implementing corporate lean programs: The effect of management control practices. *Journal of Operations Management, 36*(0), 90–102.

16.3.2. **The Quality Circle**

The Quality Circle concept was formulated by Kaoru Ishikawa in 1962.[14] Historically, it was the first *successful* attempt at engaging production operators in improvement. A quality circle is a group of workers who do similar work, who meet regularly to identify, analyze, and solve work-related problems. Figure 16.10 illustrates a Quality Circle meeting.

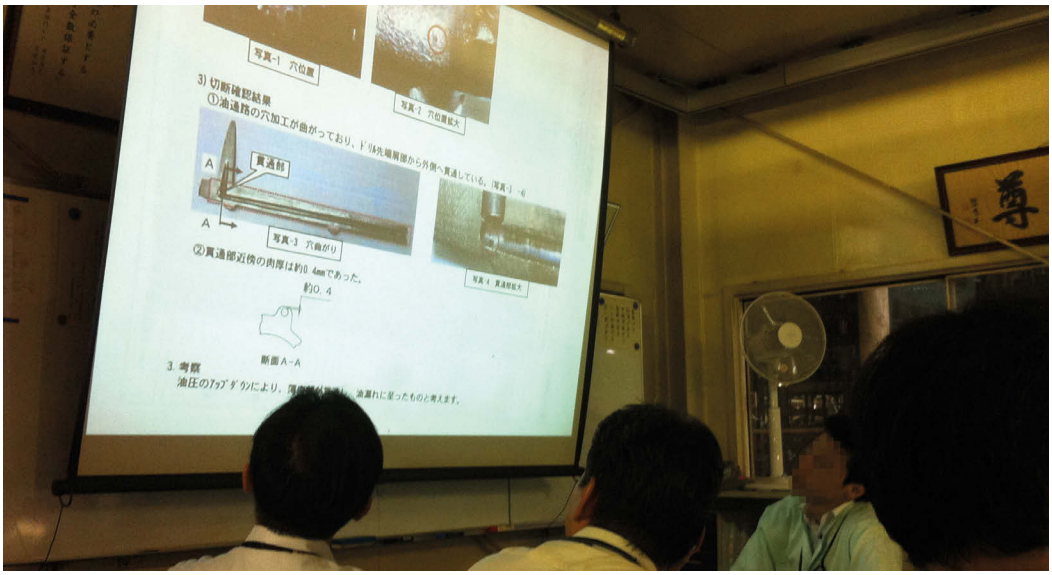


Figure 16.10 Quality circle meeting in a Japanese factory

While Quality Circles still thrive in Asia, they were a fashion that came and went in the 1980s in American and European companies. In the US, Quality Circles are found only in Japanese transplants, like Toyota's factories. The key to the success of Quality Circles was that it involved not just asking for help, but providing the "circle" as a structure, the "7 tools of quality control," and training. In its day, Ishikawa's toolbox (Box 16.4) was what made Quality Circles different from earlier attempts to organize small group activity. Instead of simply letting people loose on a problem, the toolbox gave them means to solve it.

The need for improvement in manufacturing operations, however, endures, and so does the need to involve production operators because they know process details that no one else does and the professional staff does not have the bandwidth to solve all the problems. The thinking behind Quality Circles has carried over to other types of projects. When a team starts SMED projects, it receives training on the SMED methodology; a cell project, on cell design; kanban implementation, on the rules for using kanban. In all these cases, as for Quality Circles, it is vital to put this training to use immediately.

Participation in circles is voluntary. Usually, operators sign up for a year to be part of a group of 6 to 8, meeting in overtime 2 to 3 hours per week and occasionally coming on weekends to build devices, move furniture, or run experiments. A circle takes on 3 to 4 projects per year, suggested or approved by management. The operators document and present their results in company meetings, and possibly enter their projects in company-wide competitions.

14 Ishikawa, K. (1968). *Guide to quality control.* Asian Productivity Organization.

To an uninvolved engineer, a scientist, or a medical doctor, "problem-solving" as practiced by shop floor teams may appear crude and simplistic. He or she may, for example, view a fishbone diagram as a poor excuse for a fault-tree because it makes no distinction between "OR," "XOR," or "AND" combination of causes. In the fishbone diagram, these details are not omitted for lack of sophistication but instead by due consideration of the purpose. You can fill out a useful fishbone diagram in a brainstorming session with a problem-solving team, but you would get bogged down in details if you tried to generate a fault-tree. There are many simple techniques that could potentially be applied. The value of a problem-solving method is that, for a given range of problems, it has shown itself both sophisticated enough to work and simple enough to be applied by the teams at hand.

16.3.3. Task force and the obeya

Task force is originally a military term for a temporary group under one commander to achieve a result. The members are dedicated to it while the task force exists, and return to their original units or departments once the task is accomplished and the force disbanded. The team put together for a kaizen event is a special case of a task force but task forces are assembled for tasks beyond the scope of what kaizen events can do, including, for example, creating a new plant or a new assembly line, or standardizing the dimensions of 300 dies to facilitate changeovers.

For the duration, task force members are collocated in a workspace that often includes an *obeya* or "war room." As Frederick Brooks[15] pointed out, the reason the Tower of Babel project failed is that team members could not communicate. "Obeyas" (大部屋) – project war rooms – and formal reports on a single sheet of A3 paper are communication tools developed within Toyota. Today, they are supplemented by collaboration software that has gained wide acceptance during the COVID-19 pandemic.

What is an obeya?

Obeya (大部屋) is Japanese for "big room." The term has been getting attention as a solution for project teams. The idea is to bring together in one room all the stakeholders in a project to communicate face to face, find solutions, and make decisions. It sounds like a *meeting*, and some may call their meeting rooms "obeya" without changing anything. Those who wish to dig deeper, however, find a more specific – and useful – concept, if not a panacea.

An obeya has several characteristics:

1. **It is a big room.** It may or may not have doors, and even locks depending on the need to protect the equipment or the confidentiality of the work done inside.
2. **It is dedicated.** The room is dedicated to a project team of up to, say, 10 people for the duration of the project. According to Morgan and Liker's account of Toyota's product development system,[16] the team members retain their personal desks in the functional department they belong to. If a project has multiple phases, for example from concept to production ramp-up, the obeya may migrate with it, from a location in the R&D lab to a factory.

15 Brooks, F. (1975). *The mythical man-month*. Addison-Wesley.

16 Morgan, J. & Liker, J. K. (2020). *The Toyota Product Development System: Integrating people, process, and technology*. Taylor & Francis.

3 **It contains product units.** If the team works on a product that is sufficiently small, at least one unit of the product is present in the room, with tools and measurement instruments as needed.
4 **It has standard conference room equipment.** The room has the standard equipment of a conference room, such as a white board to support informal discussions, and a projector with a screen to work on information in electronic form.
5 **Project information on the walls.** In addition, the walls are covered with status information, plans, schedules, layout drawings, and assorted charts the team finds useful.

An obeya is a big room, but not a gigantic one. Beyond facilitating communications and decision making, working in the obeya helps team members become mutually accountable as they leverage their complementary skills in pursuit of a common goal. In other words, it helps them coalesce as a team and perform at a higher level than they otherwise would. The limit of this approach is that this group dynamic is observed with groups of 5 to 10 members, but not with groups of 50 to 100.

Not every big room where multiple people work is an obeya. Contrary to the *generic* structures of office rooms, an obeya is designed *specifically* to support the activities taking place in it. Obeyas set up to develop a product, organize a conference, or recover from a natural disaster will look and operate differently. Figure 16.11, from Toyota Material Handling, shows an obeya in use. Note that the obeya contains a conference table but no desks. It should have videoconferencing opportunities to communicate with outside partners and show analyses prepared with computers. Importantly the walls are full of drawings and whiteboards, used for brainstorming, documentation, planning, and analysis.



Figure 16.11 An obeya in use (Credit: Toyota Material Handling)

Often the obeya is located as close as can be arranged to the target area and open towards it. The advantage of this approach is to keep the team at the actual place, where the actual objects of the work are (*genchi genbutsu*). It is not, however, without shortcomings:

- **Noise.** It may be noisy, making discussions and brainstorming difficult.
- **Equipment ruggedness.** Any equipment, such as cameras or computers, used by the team must be rugged enough for the shop floor environment.

- **Transparency of early or controversial ideas.** The team may be reviewing several line layout options, some of which will be poor. Seeing a poor layout on the white board while walking by the team war room may unnecessarily distress an operator who would be affected by it.
- **Security.** Securing it in the off hours in an environment that is open to hundreds of people may be a challenge.
- **Profile.** The project manager must decide how high a profile the project should have in the organization over time.

Companies *not* using obeya

There are companies that are effective at running projects, yet have no use for obeyas. For a team to meet, a project coordinator uses a computer-based calendar system to find the first time when all participants are available. He or she then reserves a conference room online and all team members are automatically notified.

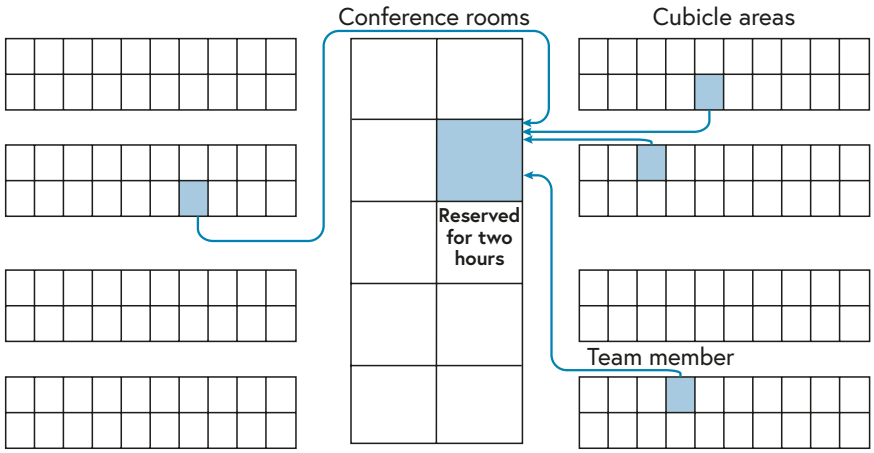


Figure 16.12 Project team meetings without obeya

In one case, when a project team requested a war room on an exception basis, the request went all the way up to a vice president who denied it. In addition, all employees in the office area are required to file away all documents and lock up their desks before leaving every day, so that no company proprietary information is left lying around. With these practices, one might expect this company's project management to be slow and inefficient but in fact it is not. Possible explanations for this include the following:

1. It is a prestigious company, able to attract top talent.
2. Time management is rigorous, so that meetings start and end on time, have agendas, and publish minutes in a standard format.
3. It is a high-technology company, with many project tasks being carried out by 1 person only.

The obeya is a tool, and management should *not* expect a team's performance to improve just because it has one. The team and its leader must learn how to use it, and *can* be expected to become better at it over time.

16.4. **Project management**

The developments of the Airbus A380 and the Boeing 787 were both more than 2 years late. Perhaps, these two master aircraft builders should know better. Such performance, however, is not the exception. Generally, from installation of a new machine to the development of a product or a production line, projects tend to be behind schedule, over budget, or both. Project management is a skill worth learning.

That it is technically possible to manage projects well – even with a large-scale project involving new technology – is illustrated by the Apollo program: It reached the goal set by president John F. Kennedy 5 months ahead of the deadline at a cost of $24B, accurately estimated in 1961. Even NASA, however, did not repeat its Apollo performance with the Space Shuttle. One hundred years before, another major project, the transcontinental railroad, was also completed in record time at a cost that was consistent with original estimates. It is therefore legitimate to ask why so many projects are late, given the tools we now have. The key is that we should not treat the challenges of project and program management as *technical* but instead as *human* and *social*. Outcomes are not determined by procedures and algorithms but by the consequences of decisions by several independent humans.

In project management, as described in the *Project Management Body of Knowledge* (PMBOK),[17] you break a project down into smaller tasks, to which you assign durations, resources, and constraints of timing or precedence. Then you run calculations that produce Gantt charts, load profiles for resources, the critical path, etc. The most commonly used tool is the *Critical Path Method* (CPM), as implemented in packages like Microsoft Project. Then there is Eli Goldratt's *Critical Chain Project Management*, focused on resources, and the *Design Structure Matrix* (DSM) to manage task iterations. These are logically consistent models but often cast aside in daily project management in favor of primitive *action-item lists*. The most obvious problem with the use of the tools is that durations for elementary tasks are not known, and most organizations do not make a serious effort at measuring them.

A project task is not like an assembly operation that you can time with a stopwatch or record on video, as project work is never completely repetitive: If you keep repeating the same action, it may be called production or transaction processing, but not a project. Since you have never done this project task before, you cannot tell *exactly* how long it will take, but you can get lower bounds based on its content – for example if it involves driving 60 miles or waiting for paint to dry – and you could estimate it based on similarities and differences with tasks you have done before, if you had kept track of them. Engineering work is organized by project, and the time of engineers is tracked through time sheets, but these sheets are filled out after the fact by the engineers themselves for bookkeeping, and are not accurate.

Most managers reject proposals that violate the laws of physics but they also reject realistic plans as "too cautious." The proposal they accept is the most aggressive schedule that is not physically impossible. The project leader knows the deadline won't be met but deals with this issue later. The manager knows it too but is betting that the project leader will redouble his efforts to atone for failure to deliver on time. The project leaders make the sale on that basis, miss the deadlines, and work desperately hard to finish. As a result of surviving in this game, they are promoted and repeat the same practices with the next generation. To change this behavior, the initiative must come from the top.

17 PMI (2021). *Guide to the Project Management Body of Knowledge (PMBOK Guide) and the standard for project management*. Project Management Institute.

16.4.1. Program Evaluation and Review Technique

In the late 1950s, Willard Fazar, Morgan Walker, and James Kelley started representing projects as networks with event milestones as nodes and tasks as preceding branches.[18] Fazar developed the Program Evaluation and Review Technique (PERT). PERT charts were first used by the US Navy's Special Projects Office in 1957 to guide the Polaris missile project. Figure 16.13 illustrates the key idea of PERT. The chart is read left to right. The final milestone is given the highest number. Numbering events with an increment of 10 allows for adding sub-events later, for example, milestone 11, 12, 21, and so on. The tasks are labeled with letters and assigned a completion time.

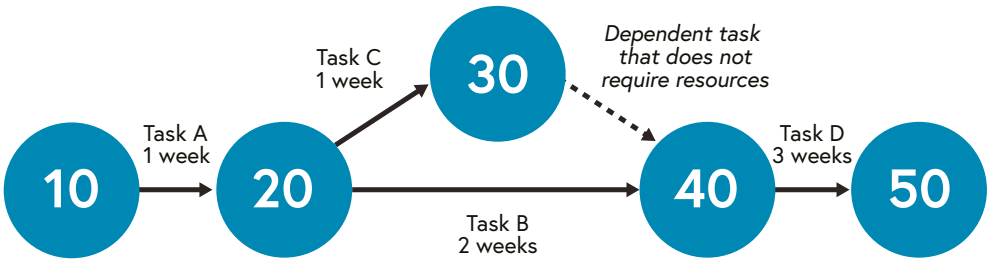


Figure 16.13 A project represented as a PERT network

PERT is focused on crossing milestones and therefore on managing times for each task, which can be deterministic or random. Because the arrows represent activities, the PERT notation can be confusing.

16.4.2. Critical Path Method

Walker and Kelley developed the CPM that is now ubiquitous in project-planning software. It only uses 1 deterministic duration for each task and identifies the sequence of tasks with no slack, known as the critical path. These are the tasks for which any delay passes onto the entire project. Conversely, any reduction in duration for any of these tasks redounds to the entire project, at least until other tasks get onto the critical path. If you work forward from start, considering precedence constraints, you find the earliest time each task can be completed; going backward from project completion, you find the latest completion time for each task. If the earliest and latest completion times are identical, the task is on the critical path. The critical path in Figure 16.14 is along tasks A, B, and D and takes 1 + 2 + 3 = 6 weeks.

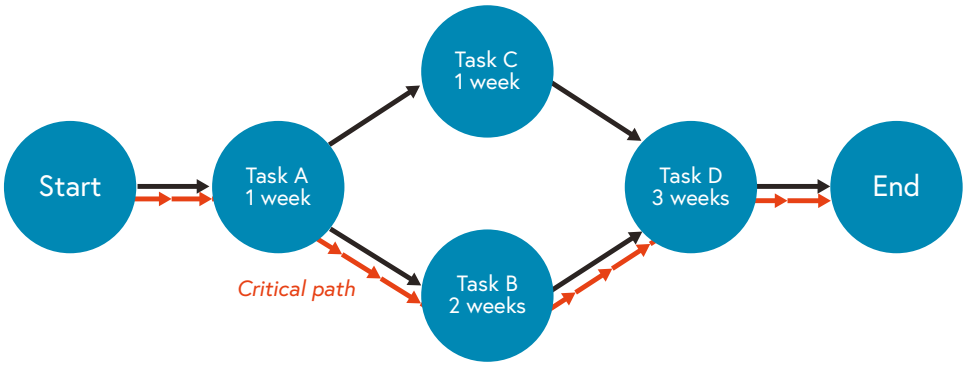


Figure 16.14 A CPM network with the critical path A-B-D

18 Kelley, J. E., Walker, M. R., & Sayer, J. S. (1989). The origins of CPM: A personal history. *PM Network, 3*(2), 7–22.

Because the critical path determines the completion date of the project, the project can be completed earlier by allocating additional resources to the activities on the critical path.

Most project management software products apply CPM. Figure 16.15 is an example of a project network diagram from MS Project. The boxes are tasks and the arrows precedence constraints, mostly finish-to-start, but some as also finish-to-finish, meaning that both tasks must be completed at the same time. The pink boxes and arrows are the critical path.

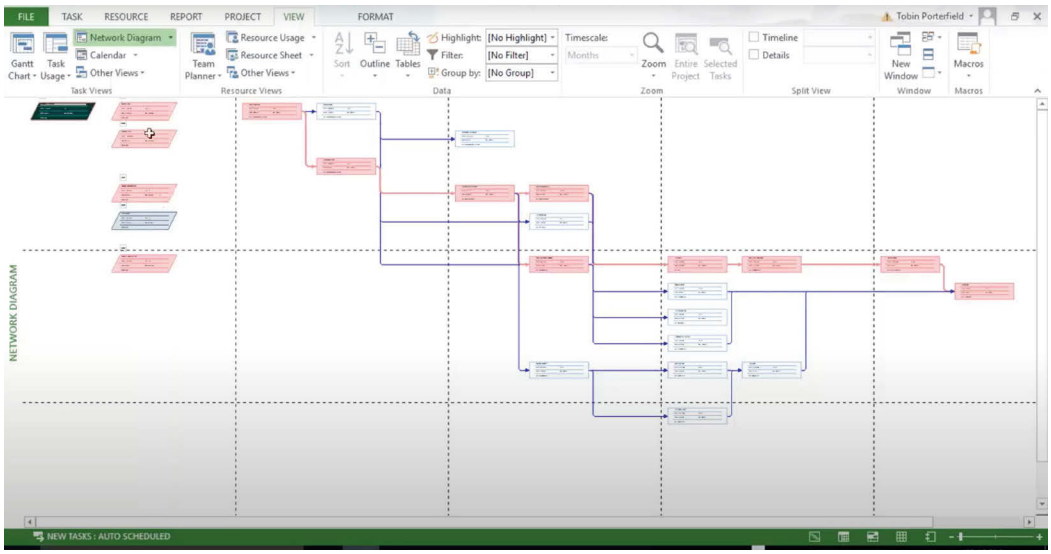


Figure 16.15 MS Project network diagram

Both PERT and CPM are *technical* approaches to a *human* problem. CPM bases its calculations on task durations that, particularly in innovative projects, no one can properly measure. The individual in charge wants task times long and the manager wants them short. They negotiate the value the system uses, and it has little to do with the time actually needed. When a project manager shows a team the critical path, it affects behaviors. Members in charge of tasks *not* on the critical path shift their priorities elsewhere and don't work on these tasks until they make it onto the critical path.

16.4.3. Critical Chain Project Management

Eli Goldratt introduced Critical Chain Project Management (CCPM) in 1997.[19] Rather than padding each task duration in anticipation of contingencies, he pads the entire project. CCPM then builds a plan with no allowance for delays on any task but provides a slack budget for the entire project, on which each task draws as needed. The project manager monitors the consumption of this budget to make it last just long enough to finish on time.

Goldratt compares this to planning a trip to the airport. If the streets and roads are empty, the airport is 30 minutes away. In reality, there will be traffic and it is prudent to budget 70 minutes for the trip rather than 30. This allowance is for the trip as a whole, not for any traffic light, freeway on-ramp, or

19 Goldratt, E. M. (1997). *Critical chain*. North River Press.

road construction site. Where the slack will be consumed is unknown ahead of time, but the budgeted amount is sufficient to arrive on time.

16.4.4. The Design Structure Matrix

Tools of the PERT/CPM generation addresses the needs of work flow with obvious precedence constraints. You have to apply paint before you let it dry. You have to build a concrete slab before you locate equipment on top of it. These project types have tasks in parallel or sequence, as shown above.

Many projects, however, have deliverables that are not hardware but information, such as the design of a product or a manufacturing process. In this context, the relationship between tasks is much more complex. The example shown here is from a car engine design project. The valve train design team cannot do its work without the output of short block system design, but the converse is also true: The short block system design team also needs the output of valve train design. The tasks are coupled. This third type of task relationship is illustrated in the rightmost column in Figure 16.16. The *Design Structure Matrix* (DSM)[20] representation is shown below.

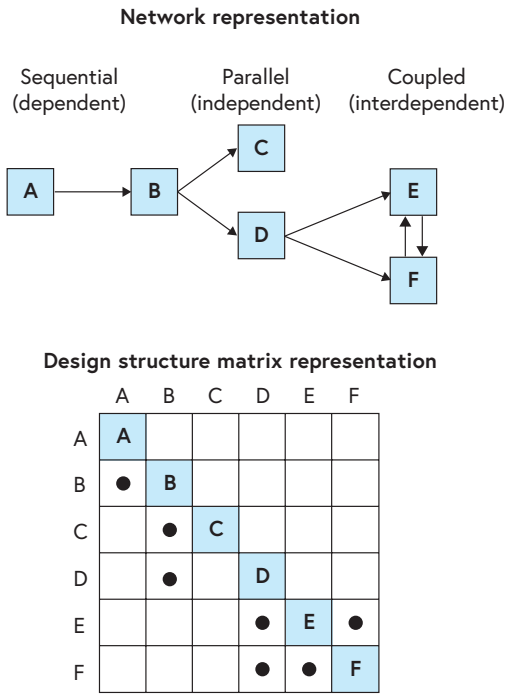


Figure 16.16 DSM patterns for two tasks

Parallel tasks are fully decoupled, in that neither one needs any input from the other. They both need to be completed for their common successors to start, but either one of them can be completed ahead of the other. The DSM for such tasks is blank. If A and B are sequential tasks, then B needs the output of A but A does not need anything from B. In Figure 16.16, the DSM shows this as a bullet where B's row

20 Browning, T. R. & Eppinger, S. D. (2016). *Design structure matrix methods and applications*. MIT Press.

meets A's column. If each task needs the output of the other, they are coupled, and this is shown in the DSM by having bullets both where A's row meets B's column and where B's row meets A's column.

Coupled tasks can only be carried out by *iterations* involving intense communications between the teams and neither one can finish ahead of the other. In the network representation of the project, this is represented by a finish-to-finish constraint. Clearly, such couplings are not desirable for the project manager. The fewer there are, the easier it is to keep control over the project, and the DSM is a tool to minimize them. You can identify coupled tasks by asking implementers in charge of each task who they need input from. The answer to this question is usually much more complete and accurate than answers to questions regarding who they provide output to.

Figure 16.17 shows an example of a DSM, adapted from Ulrich & Eppinger.[21] It only shows one DSM application, for process models, or task relationships, but there are other (static) DSM models for products, organizations, etc. The same list of tasks is shown across the top and on the left, so that the blue squares on the diagonal indicate where the row and column for the same task intersect. The Xs indicate that the row task needs input from the column task. The ×s below the diagonal represent sequential tasks, and, if all the ×s are green, the project can be represented as a PERT/CPM network. On the other hand, any × above the diagonal indicates a feedback loop with iterations. The farther the × is from the diagonal, the longer the iteration loop and the more severe its impact on the project. An o indicates an unplanned iteration.

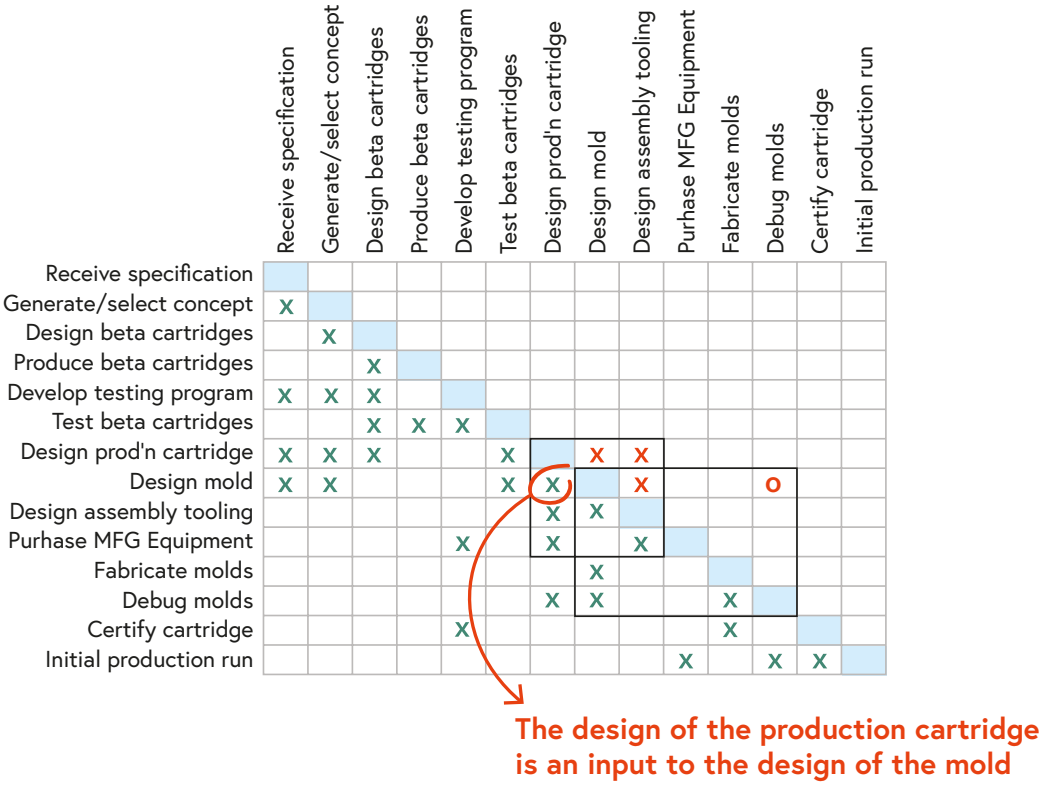


Figure 16.17 DSM for a product design process

21 Ulrich, K. T. & Eppinger, S. D. (1995). Design for manufacturing. In *Product design and development* (pp. 180–216). McGraw-Hill.

The idea of DSM is to rearrange the tasks so as to eliminate iterations if possible, or minimize their scope by bringing the red ×s closer to the diagonal otherwise. In this example, there is an unplanned loop between "Design mold" and "Debug molds." The intent of "Debug molds" is to make adjustments to molds to make them match the design better. The reality is that design errors found at "Debug molds" force mold redesigns. Plan A then is to eliminate this loop by improving the mold design process; Plan B, to find a way to fabricate and debug molds prior to designing assembly tooling and buying manufacturing equipment by using special equipment for this purpose, as is done, for example, in the production of stamping dies for car body parts.

16.4.5. Agile project management

Consultants and academics have been using the word "agile" with different meanings since the 1980s. In *project management*, the use of agile is traced to the 2001 *Agile Manifesto*[22] published by a group of software developers. The key differentiator from traditional software development methods is that agile is based on quick repeating cycles of developing and testing software functions as the project progresses. It contrasts with the earlier "waterfall" method where the whole project is planned, executed, and handed over to the customer. The agile method values routine involvement of the customer and adaptability to changing requirements and progress. It breaks down the overall project delivery to subparts that can be tested and approved sequentially. The waterfall method values long-term planning. Figure 16.18 illustrates the main difference between waterfall and agile.

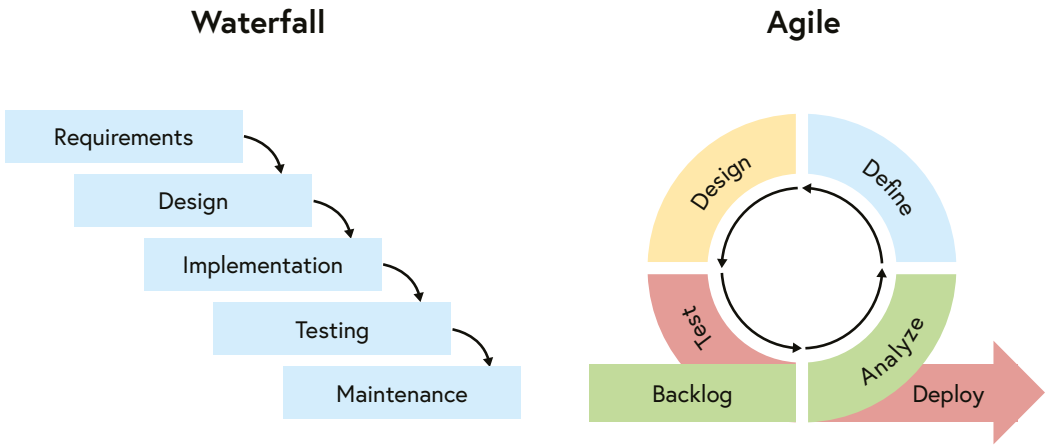


Figure 16.18 Waterfall method versus agile

The agile method uses a number of special terms and frameworks to explain how projects should be managed. A "scrum" is a team of up to 10 members that executes a software development project. The content of the project is defined through "use cases," which are single-sentence expressions of user requirements, like "Customer service agent issues advance shipping notice to customer." Each project iteration is called a "sprint." A scrum works in sprints of 10 days to 1 month, at the end of which it delivers software to customers and collects feedback. The work done during a sprint is often broken down into modules described on post-it notes on a white board with 3 columns for "to-do," "in-process," and "done" (see Figure 16.19) This is called "kanban" even though it bears no similarity to the kanban system as used in manufacturing beyond the use of a board.

22 Beck, K., Beedle, M., van Bennekum, A., et al. (2001). *The Agile Manifesto.* Agile Alliance. URL: http://agilemanifesto.org

A "scrum master" is responsible for the outcome. The scrum has committed members called *pigs*, and occasional contributors called *chickens*. This strange use of words come from the business riddle "In a bacon and egg breakfast, what's the difference between the chicken and the pig?" To which the answer is "The chicken is involved, but the pig is committed."

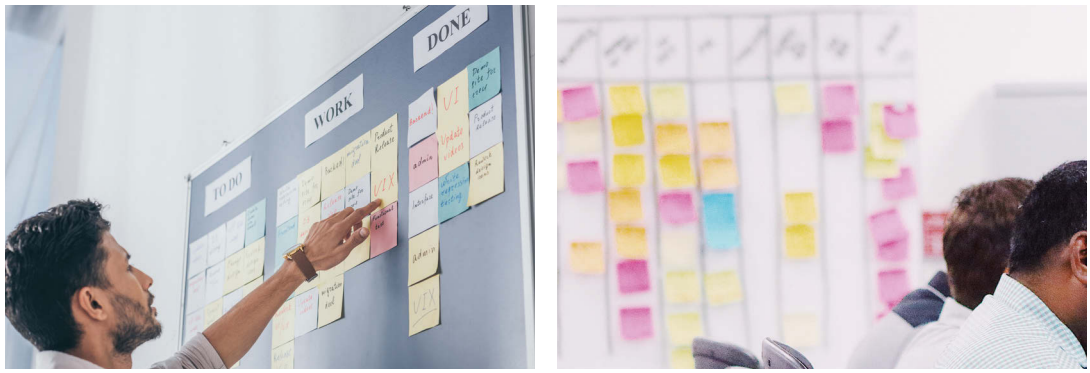


Figure 16.19 Agile "kanban" boards separate tasks into To do (backlog), Work (in-process), and Done

To drive agile project management in large corporations, Dean Leffingwell and Drew Jemilo developed the controversial Scaled Agile Framework (SAFe) in 2011. SAFe is a project management standard for implementing agile practices at scale in an enterprise. Similar to other project management standards (see Box 16.6), the company behind, Scaled Agile, is providing audits and advisory services. Companies buy into such standards because it gives them a clear framework to organize activities around, it allows them to get started quickly, and the standards have been "proven" in other firms. Support for SAFe is not unanimous among the proponents of agile.

Box 16.6 **Project management standards**

Many organizations offer standards and certifications in project management as a generic discipline. Readers may find the following frameworks of interest:

- In the US, the Project Management Institute publishes a *Project Management Body of Knowledge* (PMBOK) (1996). PMBOK is the ANSI standard.
- *ISO 10006* is a standard for project management quality from the International Standards Organization (ISO).
- *International Competence Baseline* (ICB) issued by the International Project Management Association (IPMA).
- Germany has the *DIN 69901-69905* standards for project management.
- The UK government developed *PRojects in Controlled Environments* (PRINCE2) in 1996 as a standard for information systems projects, which was later transferred to AXELOS Ltd.

Compliance with such standards is often required of government contractors, and they commonly have employees with the job title of *Project Manager* to format proposals and reports for approval by auditors. Generic project managers are otherwise not involved in the actual management of projects beyond possibly scheduling meetings, because they are not familiar with the technical content of the projects.

16.5. **Company-wide improvement models**

Many generic manufacturing models or philosophies promise superior returns and competitive advantage if only implemented diligently. New recipes commonly start as successful initiatives within a company, and solves a particular problem. Their inventors then offer it as a consulting service to other companies, usually generalizing it for a broader range of uses. When they are successful, the approach turns into a *brand* that many other consultants attach to their own offerings, thereby diluting it. As it gradually loses meaning and is oversold, interest from manufacturers wanes and the original contributions migrate into the common manufacturing body of knowledge that this book focuses on.

While students need this body of knowledge, graduates are bound to encounter a succession of brands in their professional lives and face complex decisions about their level of engagement. In particular, they may find certifications in one or more of these brands to be requirements on job applications.

Necessarily, the overlap between approaches is large and the uniqueness of each that actually has potential to improve performance may not be easy to identify. New manufacturing models – even if it is only by name – are not *only* noise and distractions. They can help renew the focus on operations, create healthy reflection and debate, and pin manufacturing improvement to the top of the management agenda. However, it is essential that companies are able to differentiate management fads[23] from substantive, actionable ideas.

In manufacturing, some management models have had proven staying power, which indicates they cannot be rejected as fads. All these models have something substantial to offer, which companies can adopt or adapt to the situation at hand. Note that there is no commonly agreed definition of any of the models; the internet is full of descriptions, opinions, and often contradictory articles on each of them. In particular, three models are worth studying:

- **Toyota Production System (TPS)**: Toyota's famous production system, on which the *lean* brand is based, discussed in more details below.
- **Total Productive Maintenance (TPM)**: A holistic model to equipment maintenance, which is particularly popular in process industries, where the term is commonly used for all improvement activities, even unrelated to maintenance.
- **Total Quality Management (TQM)**: Today's TQM is the 4th generation of the approach. In 1951, A. Feigenbaum coined Total Quality Control (TQC) as the application of quality control to all aspects of business, not just outgoing products. 20 years later Kaoru Ishikawa used the same name for involvement of every employee in quality control; TQM was coined in the US in 1986 as what the Malcolm Baldridge National Quality Award was given for. This version of TQM lost popularity in a few years but Noriaki Kano's version then gained currency in Japan.

Other notable models in manufacturing include the following ones:

- **Quick Response Manufacturing (QRM)**:[24] Dr. Rajan Suri's company-wide improvement model that focuses on time as competitive priority.
- **Six sigma**: Originally developed by Motorola in 1986, it first became famous when Jack Welsh used it to transform GE. Six sigma focuses on reducing variation and improving capabilities in

23 Abrahamson, E. (1991). Managerial fads and fashions: The diffusion and rejection of innovations. *Academy of Management Review, 16*(3), 586–612.

24 Suri, R. (1998). *Quick response manufacturing: A companywide approach to reducing lead times*. Productivity Press.

processes and produce at "six sigma" quality levels (~3.4 defects per million opportunities).[25] The statistical foundations of six sigma are questionable, but as an improvement model with tools, certification belt programs, and frameworks it is still widely used in manufacturing.

- **Business Process Re-engineering (BPR)**:[26] A blank-paper approach to improving manufacturing that focuses on mapping workflows and processes and rethinking them altogether using IT. BPR was important from the mid-1980s to mid-1990s but does not have a significant position today.
- **Agile manufacturing**: Crispin Vincenti-Brown and Ingersoll Engineers registered "Agile Enterprise" as a service mark in the mid-1980s. Their Agile Enterprise was a manufacturing company with production organized in the Ingersoll Engineers version of cells, grouped in focused factories. The term "agile manufacturing" was suggested by Roger Nagel and his team at Lehigh University in 1991[27] as a model for manufacturing in a more dynamic environment. The term has had a limited impact in manufacturing but is resurfacing today as a result of the broader agile movement from software development.
- **Industry 4.0**: The latest kid on the block is not a full-blown improvement model, but promises nothing less than a new industrial revolution. Industry 4.0 has existed as a term since it was proposed at the Hanover industrial fair in Germany in 2011. Related terms are cyber-physical production systems, smart manufacturing, and digital transformation. As could be expected, some have suggested "Industry 5.0" as the next step, oriented more towards the worker and environment than the technology-driven Industry 4.0.

The key point is that none of the improvement models are complete or sufficient for a company. Instead, companies should tailor what works for them, which means learning from a wide variety of models and best practices. This can be difficult journey, which requires living with continuous change, uncertainty, and ambiguity. To provide a structure for the improvement work, companies develop their own *company-specific production systems* (XPS) as described later in this chapter.

Some companies aim to avoid this internal system development by adopting a standard improvement model from a consulting company or certification body. This can be a good choice when there is large overlap between the models and the need, but it comes at the cost of external support and possibly less ownership and engagement. The following are examples of popular blueprint improvement models:

- **Baldrige Performance Excellence Program**: Based on the Malcolm Baldrige National Quality Award (MBNQA) established by the US Congress in 1987 to raise awareness of quality management.
- **The EFQM excellence model**: **EFQM** is for the non-profit European Foundation for Quality Management. It is an improvement model for business excellence proposed in 1992.
- **World Class Manufacturing (WCM)**: WCM was developed by Fiat and partnering firms in 2005 in cooperation with Prof. Hajime Yamashina at Kyoto Universty in Japan. It is based on TPM and consists of 10 technical and 10 managerial pillars, visualized as a temple. WCM is particularly relevant for machine shops and process industries. It is reportedly used by Chrysler, Fiat, and Barilla.
- **The Shingo Model**: Named after Shigeo Shingo, its guiding principles are enterprise alignment, continuous improvement, respect for every individual, and leading with humility. These principles are then incarnated in tools and systems and integrated into the company culture to produce results. Some companies adopt the *Shingo Model* in order to apply for the *Shingo Prize*.

25 Harry, M. J. & Schroeder, R. (2000). *Six sigma: The breakthrough management strategy revolutionizing the world's top corporations*. Currency.

26 Hammer, M. & Champy, J. (1993). *Business process reengineering*. Harper Collins.

27 Nagel, R. N. & Dove, R. (1991). *21st century manufacturing enterprise strategy: An industry-led view*. Diane Publishing.

16.5.1. The Toyota Production System and lean

Toyota holds a special position in manufacturing. Over several decades, Toyota developed its own approach to manufacturing, now known as the Toyota Production System (TPS). In 1988, John Krafcik[28] proposed to call it "lean production" to promote it outside of Toyota. Since 1988, this system has more generally been labeled "lean production," but most of what is called "lean" today bears little resemblance with the original TPS. It therefore makes sense to present TPS in its own right, and discuss lean as a *business phenomenon* or a label that – although originally based on TPS – means different things in different settings today.

TPS was a closely held trade secret until the 1970s, first formalized in an internal manual in 1973. Then, as Toyota undertook to train *suppliers* in it, word of the existence of TPS and some of the tools leaked out to Toyota competitors and to the world at large. The first authorized statement about TPS in Japanese was Taiichi Ohno's 1978 book.[29] The first original English-language books covering TPS were Robert E. Cole's analysis of HR practices in 1979,[30] Richard Schonberger's overview in 1982,[31] Robert W. Hall's detailed account of the system in 1983,[32] and Kiyoshi Suzaki's introduction in 1987.[33] At the same time, Norman Bodek started Productivity Press and started publishing translations of Japanese books, including books of Shigeo Shingo, Yasuhiro Monden, Kenichi Sekine, and the JMA, among others.

The literature on TPS is large and opinions differ. What is sure is that the TPS was developed gradually over a long period and had many originators and influences, including leading members of the Toyoda family. Of particular importance is the role of Taiichi Ohno – sometimes called the father of TPS – who together with colleagues gradually developed a production system outperforming all competitors at its time. The English translation of his 1978 book *Toyota Production System* appeared in 1988, the same year as Yasuhiro Monden's book with the same title.[34]

International attention to TPS was boosted by the *International Motor Vehicle Program* (IMVP, 1980–2013), an international benchmarking study of global automobile production funded by the Sloan Foundation and hosted by MIT. In 2013, the IMVP was reorganized into the Program on Vehicle and Mobility Innovation (PVMI) hosted by the Wharton School at the University of Pennsylvania. A common explanation in the 1980s was that Toyota's performance could be explained by the unique Japanese work ethic and culture. But the research team proved this wrong, and used among other evidence the NUMMI plant in Fremont, California (1984–2010) – a joint venture between General Motors and Toyota – to show that Toyota's model could be transferred overseas.

In 1990, James Womack, Daniel Jones, and Daniel Roos published the *The Machine that Changed the World*,[35] which summarized the IMVP study, became a bestseller, and popularized lean far beyond the

28 Krafcik, J. F. (1988). Triumph of the lean production system. *Sloan Management Review, 30*(1), 41–51.

29 Ohno, T. (1988). *Toyota Production System: Beyond large-scale production*. Productivity Press.

30 Cole, R. E. (2018). *Work, mobility, and participation.* University of California Press.

31 Schonberger, R. (1982). *Japanese manufacturing techniques: Nine hidden lessons in simplicity*. Free Press.

32 Hall, R. W. (1983). *Zero inventories*. Dow Jones-Irwin.

33 Suzaki, K. (1987). *New manufacturing challenge: Techniques for Continuous Improvement*. Free Press.

34 Monden, Y. (2010). *Toyota production system: An integrated approach to just-in-time*. Productivity Press.

35 Womack, J. P., Jones, D. T., & Roos, D. (1990). *The machine that changed the world.* Rawson Associates.

automobile industry. In the book, the authors wrote that Toyota's system was way better than the mass production methods used in the west on almost all performance categories:

> *[TPS] is lean because it uses less of everything compared with mass production – half the human effort in the factory, half the manufacturing space, half the investment in tools, half the engineering hours, to develop a new product in half the time. Also it requires keeping far less than half the needed inventory on site, results in many fewer defects, and produces a greater and ever growing variety of products.*
>
> Womack et al., 1990, p. 11

According to Daniel Jones, the first version of the TPS "house" (one version can be seen in Box 2.3 on page 52) was developed by Fujio Cho, and Ohno's team distilled it to teach TPS to suppliers. It was later used by John Shook in training materials for NUMMI.[36] Many authors have published different versions of TPS houses. Most of them, however, share these following fundamental characteristics:

- **The roof is the objective**, creating customer value by producing at the highest quality, lowest cost, and fastest delivery times.
- **The foundation is operational stability**, achieved through means like standardization, balanced workload, and removal of waste.
- **In-between the foundation and the roof are two pillars**, *just-in-time* and *jidoka*.

More details are often added, with the names of tools as part of the foundation, columns, or roof, and it is where the published house diagrams diverge.

The house metaphor is visually attractive and provides a high-level view of the roles of some components of TPS. It doesn't, however, show how these components interact and it's not an implementation road map. If you needed everything in the foundation to be in place before you start erecting the columns, and if the columns were to be in place before putting on the roof, the transformation of a factory would be practically impossible, as no management would have the patience to wait for the years it would take before reaping any tangible result.

A brief history of the development of the TPS

To understand TPS deeply, you need to work for Toyota for 10 years. Short of this, you can study it. There are many good books and reference works on TPS, including those of its founders themselves, listed in this chapter's Further Reading list. This section gives a succinct introduction.

As Takahiro Fujimoto[37] explains it, the people of Toyota found solutions to overcome crises throughout the life of the company, which eventually coalesced into a system. The techniques they adopted from others, modified, or invented are easiest to understand within their historical context. On its 75th anniversary, Toyota published a detailed timeline of the development of TPS from 1945 to 2005,[38] highlighting the key challenges the company faced in each period, and the solutions it adopted in just-in-time and jidoka. Each item has a short explanation in text, and is illustrated with cartoons, technical drawings, and photographs. It is an account of the technical content of TPS, and shows how the

36 Narusawa, T, & Shook, J. (2009). *Kaizen express: Fundamentals for your lean journey*. Lean Enterprise Institute.

37 Fujimoto, T. (1999). *The evolution of a manufacturing system at Toyota*. Oxford University Press.

38 URL: http://www.toyota-global.com/company/history_of_toyota/75years/

pieces fit together. Based on this timeline, other details contained in the 75th anniversary website, and other sources, Figure 16.20 summarizes the development of TPS, going back farther in time, and emphasizing international exchanges.

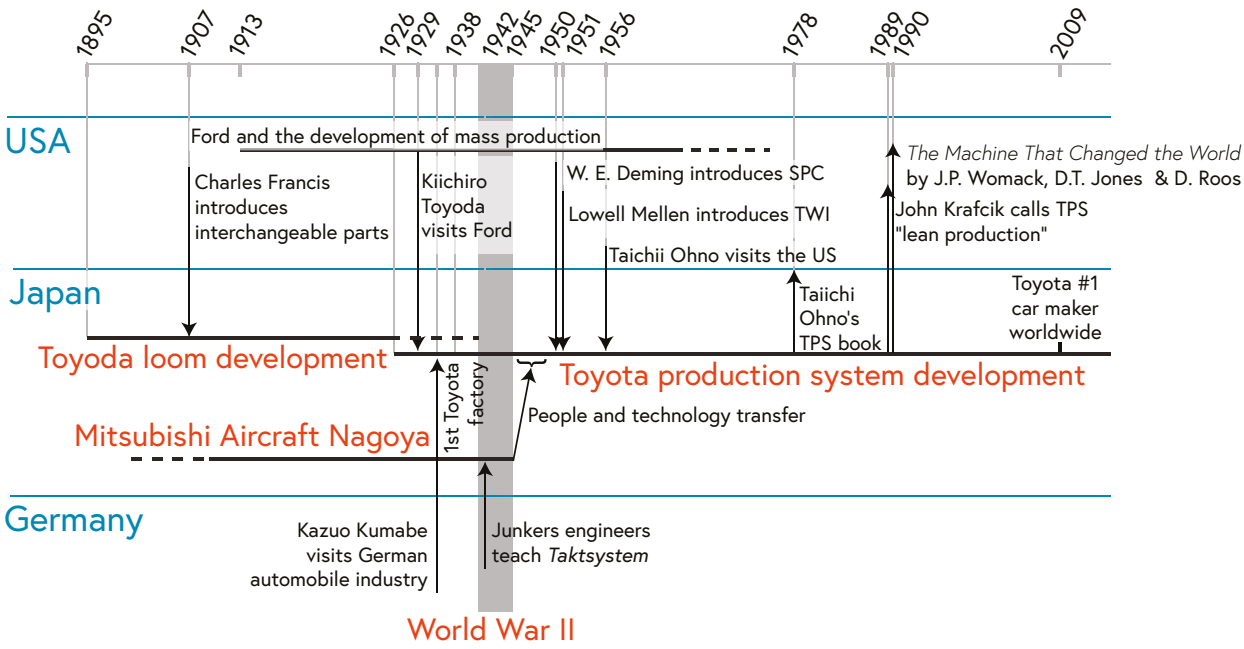


Figure 16.20 TPS development timeline

The foreign inputs to TPS, primarily from the US and Germany, were over by the mid-1950s, and, since the late 1970s, the flow has been in the opposite direction, with the world learning from Toyota. TPS is still a work in progress. It has been and still is primarily an original development. The bulk of TPS has come from the minds of inventor Sakichi Toyoda, his son Kiichiro, engineers Taiichi Ohno and Shigeo Shingo, and hundreds of thousands of other Toyota employees over decades.

Before World War II, Toyota had studied automotive technology and product design from Germany, which included reverse engineering a 1936 DKW. Toyota's first postwar model, the 1947 SA, looked like a Volkswagen Beetle. For manufacturing, on the other hand, they had turned to the US, where the car industry was five times larger than Germany's. After the war, Toyota recruited alumni of the former Mitsubishi Aircraft plant in Nagoya, who had learned the "Taktsystem" from German aircraft manufacturer Junkers.

The two key pillars of the Toyota Production System – just-in-time and jidoka – were developed by members of the Toyoda family. Before cars, they made looms. Sakichi Toyoda, the founder of the loom business, is honored in Japan as a great inventor. He spent 30 years improving the loom, one of the oldest machines known to humans. Step by step, he went from the wooden, manual loom his mother operated in the 1890s to the Type G, a powered loom with automatic shuttle change that stopped whenever the yarn broke, in 1924. Sakichi Toyoda's approach became the basis for Toyota's jidoka.

Later, Kiichiro Toyoda developed the idea of just-in-time after Toyota had started producing cars. Toyota did not have excess of space and money at that time and needed to find a way to produce cars

with minimum waste. They started with two used car presses they bought from the US and went to study automobile production around the world. Post World War II, cash-strapped Toyota could only afford to buy the quantities of materials on cars it had sold, which gave rise to the kanban system. In 1950, Kiichiro Toyoda laid off 40% of Toyota's workforce to avoid bankruptcy, and resigned to take responsibility. His cousin Eiji Toyoda took over and led the company through decades of growth.

Ohno and others ended up developing a production system on a foundation of stable processes with orders of magnitude less inventory than their competitors and consequently shorter lead times and better quality, approximating takt-driven production.

Box 16.7 **Development of SMED in Toyota**

Industrial engineer Taiichi Ohno, who had joined Toyoda Spinning in 1932, joined the Toyota Motor Company in 1943 as a shop floor supervisor. Over many years, Taiichi Ohno experimented with quick changeover of dies of the car presses and got it down from 1 day to 3 minutes. Instead of using changeover specialists, which was the standard prior practice, the die exchange could now be done by the operators themselves. This invention made it cheaper to run small batches than the large batches used by competitors. It enabled just-in-time production and eliminated the need to carry large expensive inventories, but more importantly it made stamping mistakes almost immediately visible. This enabled Toyota to produce higher quality cars at a lower cost and faster then its competitors. Later, Shigeo Shingo wrote about the technique, calling it *single-minutes exchange of die* (SMED),[39] where "single-minutes" meant "under 10 minutes."

Stop and Think 16.2
How did shorter changeover times reduce lead times and improve quality?

The leaders of Toyota knew that the only way to develop and maintain such a system was to engage employees at all levels. The employees had to be disciplined and intelligent to deal with any variation in production. They also had to contribute knowledge on how to continuously improve the processes. People are therefore at the center of the TPS, which sometimes is called the "*Thinking People System*" instead of the Toyota Production System.

TPS is often falsely interpreted as a system for the factory shop floor *only*. Although Ohno recommended to start TPS at the final assembly line and work backwards through the production system, it doesn't mean he saw it as system only to be implemented in production. From the start, Toyota noticed that the just-in-time principles only worked when the whole supply chain was aligned. Excessive fluctuations in demand or poor-quality deliveries from suppliers broke down the just-in-time system. Toyota also noticed the need to design products for efficient, mistake-proof assembly.

Toyota also took steps to manage the demand side closely via its dealer network. Toyota gradually stopped building cars in advance for unknown buyers and started instead to build and deliver cars that were already sold by a dealer to a customer. The dealers followed up with the end-customer, and did

39 Shingo, S. (1985). *A revolution in manufacturing: The SMED system*. CRC Press.

home visits to help families plan their next car. Today, while door-to-door car sales are long gone, Toyota suppliers do hold new cars for sale but, at least in Japan, Toyota seeks to stay in touch with customers through other channels, such as, for example, the gazoo.com web service it operates.

Toyota also tightly integrated its supply chain. Since the 1970s, they have been working with local suppliers and helping them in their capability improvement through supplier network programs. Rather than the adversarial supplier relations common in the US, Toyota collaborated openly with suppliers to make them as reliable and productive as possible.

In sum, TPS is a production strategy that differs radically from what the competitors did before the 1990s. Meanwhile, competitors have studied TPS and learned the related principles, techniques, and tools and implemented many of them in their factories. In the automobile industry, TPS isn't the same formidable competitive advantage it used to be, but, on many performance dimensions, Toyota still outperforms their competitors. Why? Because there is much more to TPS than tools and techniques. Peter Hines and colleagues at the Lean Enterprise Research Centre captured this point using the analogy of an iceberg.[40]



Figure 16.21 There is more to TPS than tools and techniques: The lean iceberg model of Hines et al. (2008, p. 9)

Although the underwater part of the iceberg is more tacit and harder to access, there are plenty of books and reports that seek to capture it. An example is *Toyota Culture* by Jeff Liker and Michael Hoseus (2008).[41] Toyota itself has published a manual on the culture underlying TPS for its own employees, as explained next.

40 Hines, P., Found, P., Griffiths, G., & Harrison, R. (2008). *Staying lean: Thriving not just surviving*. Lean Enterprise Center, Cardiff University.

41 Liker, J. K. & Hoseus, M. (2008). *Toyota culture*. McGraw-Hill Publishing.

The Toyota Way

When Fujio Cho became President of Toyota in 1999 he wanted to reinforce the TPS culture globally by making the behavioral aspects of TPS explicit. In 2001, Toyota published an internal document called *The Toyota Way*.

The Toyota Way is based on a foundation of five leadership behaviors:

- **Challenge.** Do not be satisfied with status quo.
- **Kaizen.** Always look for a better way.
- **Genchi genbutsu.** Means "go and see" what happens where the work is done.
- **Respect.** Show respect for what machines and humans can and cannot do.
- **Teamwork.** The best results are achieved when people work together.

On this foundation, Toyota emphasizes two cultural pillars:

- **Wisdom and kaizen.** Even Toyota translates the original chie to kaizen (知恵と改善), as "continuous improvement," leaving out the part about chie (wisdom).
- **Respect for humanity.** This is usually translated to the misleading "respect for people."

Both wisdom and kaizen have already been covered earlier in this chapter. "Respect for humanity" is not about being courteous but about taking full advantage of the potential employees have for being human, from physical strength to the ability to sense anomalies, learn, and create solutions. The practice of respect for humanity includes common-sense behaviors like acknowledging good work and caring for employees, but it also includes two concrete actions in human resource management that are often not described in the literature:

1. **Provide stable jobs.** Employees' commitment to continuous improvement and the accumulation of unique know-how cannot be achieved if they know they will be laid off at the slightest economic storm. Brain drain cannot be avoided if employees are not given career prospects.
2. **Provide career plans to employees at all levels.** This requires a strengthened Human Resources department that identifies the ambitions of each employee, including operators, and plans with each the means to achieve them through job rotations, training, and participation in improvement projects. It is unusual at the level of operators in most firms, but not at Toyota.

From TPS to lean

There are many ways to look at, define, and interpret lean production, manufacturing, or management. Many authors and practitioners disagree on definitions and spend lots of energy in fighting for camps and viewpoints of "what is lean and not." It is usually a waste of intellectual capacity and time.

The following are among the most common definitions or understanding of lean:

- **Lean as TPS.** Two usual definitions is that lean production "is TPS" or "is TPS applied outside Toyota." The problem with this definition is that TPS continuously evolves and TPS is the best fit for Toyota. When a company in the semiconductor industries, petrochemical industries, or toy industry says it seeks to implement lean as part of its production strategy, does it then mean a blueprint of TPS? Obviously not, because it would be a mismatch between its needs and strategy.

- **Lean as principles.** In an attempt to help manufacturers pursue lean production, Womack and Jones wrote a follow-up to *The Machine*, titled *Lean Thinking* (1996).[42] In this book, they introduce five lean principles that are often cited. However, these principles are at a level of abstraction that most managers can claim to be applying them:
  - Define customer value
  - Identify the value stream
  - Create flow
  - Implement pull production
  - Continuously improve the processes
- **Lean as practices.** Others have sought to understand lean as a socio-technical system that is represented by a set of practices. Rachna Shah and Peter Ward have written the two all-time most-cited academic articles on lean,[43] which operationalize lean as a list of practices sorted into bundles. Their 2003 article lists four bundles of firm-internal practices: Just-in-time, lean-related human resource management, Total Quality Management, and Total Productive Maintenance. Their 2007 article separates supplier-related, customer-related and internally-related lean practices as shown in Figure 16.22.

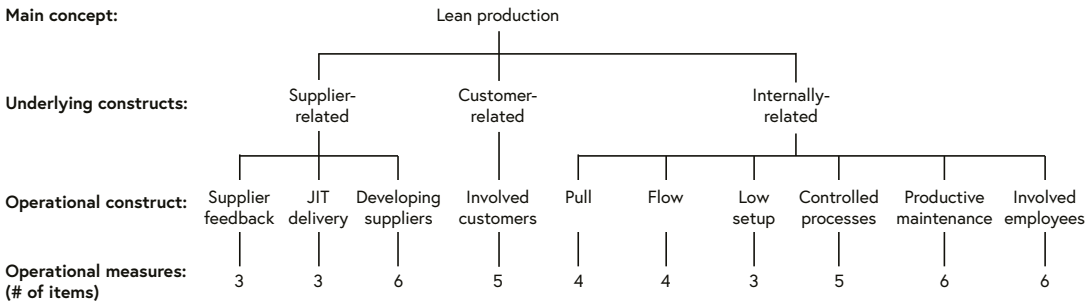


Figure 16.22 Lean understood as 10 operational constructs (Source: Shah and Ward, 2007, p. 799)

- **Lean as a business phenomenon.** With lean spreading and getting watered down over the past decades, what is left is the phenomenon. Lean has become a label used to rally people around almost any kind of improvement. This doesn't mean it cannot have a specific meaning in a specific setting. In fact, many companies develop their own "lean programs," or company-specific production systems, which gives lean a specific meaning for them.[44]

Stop and Think 16.3
Why do you think the IMVP research team chose the term "lean"?

42 Womack, J. P. & Jones, D. T. (1996). *Lean thinking: Banish waste and create wealth in your corporation.* Free Press.

43 Shah, R. & Ward, P. T. (2003). Lean manufacturing: Context, practice bundles, and performance. *Journal of Operations Management, 21*(2), 129-149. / Shah, R. & Ward, P. T. (2007) Defining and developing measures of lean production. *Journal of Operations Management, 25*(4), 785–805.

44 Cusumano, M. A., Holweg, M., Howell, J., Netland, T., Shah, R., Shook, J., Ward, P., & Womack, J. (2021). Commentaries on "The lenses of lean". *Journal of Operations Management, 67*(5), 627-639.

16.5.2. **Company-specific production systems**

Taking inspiration from Toyota's success with its Toyota Production System, firms have developed their own company-specific production systems, or XPS for short, where the X stands for the company's name and PS for "production system."[45] Some companies use alternative names like "business system," "performance system," "excellence system," "operations system," or other labels. Since the early 1990s, the development and deployment of XPSs has been an ongoing trend as shown in Figure 16.23. Today, almost all global manufacturers in a wide variety of industries have official improvement systems that resembles an XPS. For example, some known examples are the Nissan Way, Danaher Business System, Scania Production System, Bosch Production System, Alcoa Business System, Electrolux Manufacturing System, and Volvo Production System.

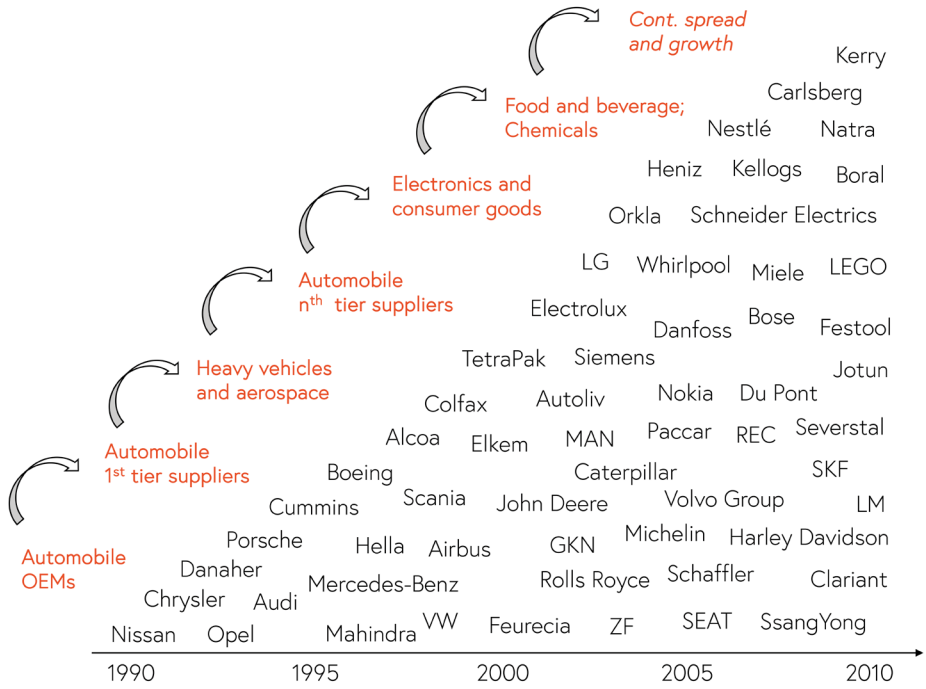


Figure 16.23 Approximate years for launching each firm's company-specific production systems (Source: Authors)

As can be seen in Figure 16.23, the trend of developing XPSs has spread from car makers to their suppliers and suppliers' suppliers in the process industries. It also spread to heavy vehicles and aircrafts, medical industries, furniture, household appliances, and food industries. Today, even service companies, hospitals, and public government agencies have developed XPSs.

Most of these systems are largely modeled on lean and the Toyota Production System, and therefore many companies often refer to their systems as their "lean programs." However, note that company-specific production systems are intended to be *tailored* to the firm's need and context. Not all lean principles and practices fit all companies and settings, and there are many principles and practices outside Toyota that can be helpful for improving the performance of a firm.

45 Netland, T. (2013). Exploring the phenomenon of company-specific production systems: One-best-way or own-best-way? *International Journal of Production Research, 51*(4), 1084–1097.

The "X-factor" of an XPS is the adaptation of the improvement model to the specific conditions and needs of a company. The use of the company's name and corporate design shows that the XPS is "ours," in the same way that the TPS is Toyota's system. XPSs are often summarized in visual models. In the automotive industry, these models often resemble the TPS house closely, but in other industries the models can take completely different forms and include other principles and practices. For example, Figure 16.24 shows the visual model of Electrolux Manufacturing System, as photographed in a factory producing washing machines.

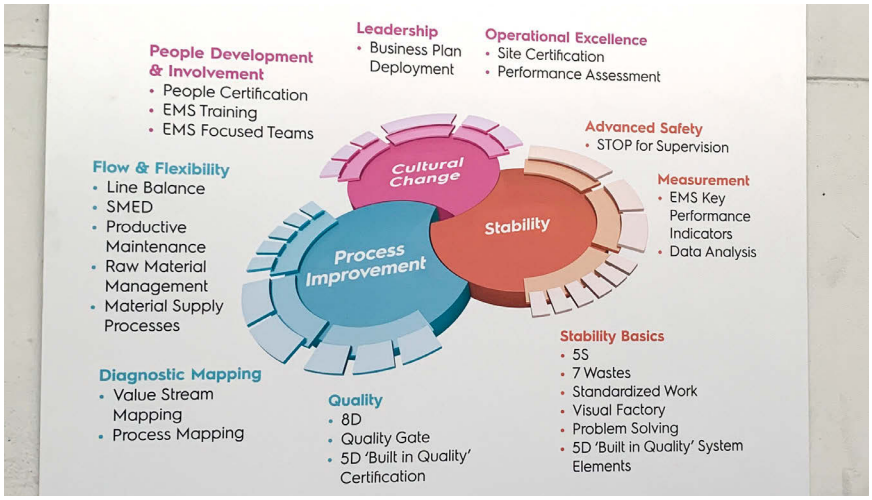


Figure 16.24 A poster of the Electrolux Manufacturing System in an Electrolux factory

XPS models typically follow the design structure exemplified in the Electrolux Manufacturing System: *Objectives* are supported by *principles* supported by *methods and tools*. In Electrolux, the objectives are "cultural change" for the better, "process improvement," and "stability." For each objective, three principles are defined, and for each principle a bullet list summarizes useful methods and tools.

Developing an own production system is a strategic approach to improvement, which supports the sharing of best practices, creates a common language for improvement work, and helps develop management commitment to continuous improvement. It helps communicate what managers are supposed to drive and what employees are supposed to do. Importantly, it differs from improvement projects in its intention of being permanent.

There are, however, plenty of pitfalls. According to reports, at least two out of three such programs produce disappointing results or fail.[46] Having an XPS is only the first step towards creating systemic improvement. Nothing improves if all there is to it is posting illustrations on the shop floor and talking about it in meetings and reports. Unfortunately, such "window dressing" is a common behavior. Many firms experience the following problematic challenges:

1. **Seeing XPS as a goal.** When the XPS model is treated as the target instead of operational improvement, there will be lots of activity and bureaucracy but little improvement. The XPS model should only provides possible guidance for achieving the target of improvement.
2. **A production-only model.** Partly due to its name, the "production system" often never makes it beyond production. When this happen, the XPS degrades to a model used in part of the factory,

46 Pay, R. (2008, March 1). Everybody's jumping on the lean bandwagon, but many are being taken for a ride. *Industry Week*.

but nowhere else.

3 **Focus on methods and tools.** The focus on codification in an XPS favors the easily modifiable, like tools and methods, while the more important soft part of improvement, like leadership and human motivation, is taken lightly.

4 **Lack of commitment.** As for any program, the support and commitment to XPS vanishes fast as the "newness" factor disappears and other initiatives start competing for management attention, like any of the above-mentioned improvement models. Without ownership in the top management group, no XPS will survive the first few years.

16.5.3. Digital transformation and smart factory programs

As a consequence of the so-called 4th Industrial Revolution[47] or "Industry 4.0," companies look to digital technologies for the next productivity leap. Over the past decade, it has become commonplace to launch "digital transformation programs," "smart factory programs," or the like.

These programs share many characteristics with lean transformation programs or XPS deployments in the way they are managed and organized but there are differences:

- **Technology-focused.** The production improvement methods discussed so far start with the processes and people managing them. "Digital transformation" programs start with technology, not needs. The ultimate objective is similar to the visions of Computer-Integrated Manufacturing (CIM) from the 1980s, with the difference that today's technology is more advanced.
- **Use case-driven.** The starting point for most applications are a "use case," often with uncertain customer value and economic value. Companies build portfolios and libraries of "use cases."
- **Broader business focus.** Even though Industry 4.0 is about manufacturing, IT has had a stronger impact on other activities. At the center of discussion is business model transformation rather than operational improvement. George Westerman, Didier Bonnet, and Andrew McAfee separate three main areas of digital transformation, of which classic process improvement is the third:[48]
  - Business models
  - Customer experience
  - Operational processes
- **Non-manufacturing reference companies.** While lean has Toyota and other showcase companies in manufacturing, the "digital" leaders are Amazon, Netflix, Uber, Airbnb, Google, and Facebook. Some manufacturing companies, frequently labeled Industry 4.0 lighthouses are not only users but also providers of the technology, like Siemens, ABB, Schneider Electric, and Bosch, which may bias their cases.
- **Requires technology expertise.** Many technologies for manufacturers under the "Industry 4.0" umbrella require specialized expertise. Consider, for example, AI, blockchain, 3D printing, Augmented Reality, cobots, or IIoT. These technologies are arguably more challenging to teach to the entire workforce than the principles of lean.

Examples like Spotify, Airbnb, or Uber make it clear the IT and OT have indeed enabled new ways of doing business. Manufacturers must engage in technology watch and experimentation, and sand-boxing of new technologies allows them to stay up to date with the newest, potentially disruptive technologies.

47 Schwab, K. (2016). *The fourth industrial revolution*. World Economic Forum, Geneva, Switzerland.

48 Westerman, G., Bonnet, D., & McAfee, A. (2014). The nine elements of digital transformation. *MIT Sloan Management Review, 55(3), 1–6.*

It can be overdone, leading the organization to spend so much effort on learning new technology that it has none left to put them to use, a disease Alvin Toffler called "future shock."[49]

Which technologies to consider and test is a question that is increasingly difficult to answer. To capture new external technologies, manufactures have to engage in "open innovation;" the ability to source innovations externally.[50] Oddly, the most determined resistance to IT innovation is found in IT departments. There are also many "lean Luddites," who base their hostility to IT and OT on writings on TPS from the 1970s and 1980s. The same purists, however, post their skepticism on social media and communicate via smartphone apps... Doing IT and OT well can be a formidable competitive advantage, and also Toyota has not been shy about using them.

For Industry 4.0 or smart factory programs, manufacturers become more dependent on outside suppliers for software, data storage, and cybersecurity, and often need to integrate services from multiple suppliers of various technologies. There is a lock-in danger with each supplier because of costs and risks of switching to another. In addition, manufacturers need help from technology consultants. To co-develop new solutions they team up with startups, vendors, and universities. All these activities can help improve productivity, or can drain resources and energy away from kaizen.

Figure 16.25 shows two men in a warehouse with a computer and VR glasses. What they do could be totally unrelated to current and future shop floor performance, or it could be a new and useful warehouse planning systems. We cannot tell without the details that are not in the picture.



Figure 16.25 Typical illustration of digital transformation in a factory

Stop and Think 16.4

If a company is running both a lean program and digital transformation program, should they merge them? What are the pros and cons?

49 Toffler, A. (1970). *Future shock*. Random House Publishing Group.

50 Chesbrough, H. W. (2006). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.

16.6. Leading a transformation program

Even more than individual improvement projects, the transformation of a company is a move into uncharted territory. No one at the outset has the information needed to establish a two-year roadmap. Management, at the outset, doesn't know enough about TPS or continuous improvement, and consultants do not know enough about the business of the company. The long-term goals can only be generic, like "move to takt-driven production," and gradually come into sharper focus as the organization learns from successful and not-so-successful projects.

Initially, you can expect support from a minority of employees. The majority will be hesitant, and there will be a minority of antagonists, rooting for the effort to fail. Announcements, exhortations, and threats do not change minds. What does is working below the radar with supporters to achieve initial successes, publicize them to win over the fence-sitters, and thereby overwhelm the antagonists. The following strategy has proven effective:

1. **Start with depth-first pilot projects.** A small number of pilot projects transform a segment of a process led by a supportive first-line manager. They achieve tangible results as an increment of movement towards the goal, that start the acquisition of skills in the organization and motivate other projects.
2. **Engage managers from the top down.** Giving pep talks at kick-off meetings, regularly visiting teams during projects, and attending report-outs is a start. When projects proliferate in a plant, management sets up a steering committee to vet projects, coordinate resources, and remove road blocks as needed.
3. **Engage production control and logistics.** Once improvements in production itself have opened a visible gap with the established means used to purchase, move, and deliver materials to lines, projects are initiated in these areas to close this gap.
4. **Engage the support departments.** Support departments like Maintenance, Quality Assurance, and Human Resources then adjust their structure, procedures, and practices to support the improved productions and logistics processes.
5. **Initiate Hoshin planning** **(See Box 15.4 on page 599)**. As projects proliferate, employees start questioning what they add up to and where the plant is heading. It is the signal that Hoshin planning is needed to provide a direction for the next 1 to 3 years.
6. **Start "Total" activities.** "Total" is used in Japan to mean "with participation by everyone." 5S, TQM, and TPM are examples of company-wide activities with *mandatory* participation, and you cannot issue such mandates until the workforce is ready. This happens once you have substantial results based on *guided volunteerism*, in which engagement is voluntary but management clearly encourages and welcomes it. Once guided volunteerism has lifted workforce engagement from 0% to ~20%, you can leverage the engaged minority to bring along the others.

The challenge of achieving and sustaining 100% participation requires years of sustained effort. "100%" includes managers. 5S and TPM entail cleaning machines, and the workforce needs to see the plant manager don a jumpsuit and personally participate in cleaning a machine. Otherwise, it's a mixed message.

The "Total" activities are *breadth-first*, in that they affect all sections, departments, and areas. It makes their effect difficult to quantify, sometimes even once in full swing. It does not mean they are ineffective but it does mean that they require a greater leap of faith from participants than projects that are justified upfront by reductions in lead time and WIP, and productivity increases.

The "Total" activities require training at all levels, starting from top management. All members of the organization receive training focused on their specific roles in promoting the activities within their area of responsibility, supporting them on an ongoing basis, and personally participating. Until the activities become integrated in "the way we work around here," management must regularly check participation.

16.6.1. Starting from the top

Some managers' reaction to the challenge of improvement is that their own operations are already run as effectively and efficiently as possible, given external constraints, including fickle customers, unreliable suppliers, local labor laws, worn-out equipment, a rickety transportation infrastructure, and the weather. They see internal problems as so puny compared to the external ones that it would be futile to address them. "Who cares about doubling operator productivity when labor costs are 5% of the cost of goods sold?" they ask, or "Why should we reduce production lead times to one day when component purchasing lead times are four months?"

The gist of all these reactions is that the causes of problems are all external and the solutions are for *others* to change. The managers are saying, "My own organization is optimal, and performance cannot improve unless my suppliers, customers, and complementers get their act together." Such statements, however plausible, are never true, as there is no such thing as operations that cannot be improved. Even if there were, the question would remain of how you would persuade *others* to improve.

The initiative for improvement *must* come from management, with the top decision maker sufficiently involved to make his or her commitment obvious to all. The impetus must come from the top. In companies with multiple facilities, plant managers usually do not have enough authority. It is also for management to decide when implementation has proceeded far enough to engage partner organizations like suppliers. By that time, there usually is a steering committee of upper management to review that decision.

16.6.2. Letting the XPS emerge

Given the lack of knowledge at the beginning of the transformation, it is a mistake to mandate tools, procedures, or formats upfront. If, within broad guidelines, you let plant staff be creative in their own projects and organize the sharing of results, you are allowing a new system to *emerge*.

A medical devices company started its transformation in 1984. Nine years later, they compiled documents from actual projects in multiple plants into an internal how-to book with an introduction by the CEO. It was a form of an early XPS. Every new engineer or manager received a copy. They revised the book in 2010. The company grew 10-fold in size during that period, under steady, patient leadership.

An automobile company developed an XPS in 2007. Seven years later they had learned a lot from seeking to implement it, were further in the journey, and got a new top management team. It was time for the company to renew, rename, and rethink the XPS, including where it was governed. The new XPS launched in 2015 provided a needed boost to the improvement culture and allowed more local ownership.

Different units and factories in a network will take different positions towards a corporate improvement program:[51] Avoid, act, adopt, or adapt (see the 4A framework in Box 16.8). A primary job of the central improvement team is to help units move towards the right in the framework. Adapting the corporate blueprint to local conditions is the best way, but it may require the local site to first adopt the global standard to learn the underlying practices. As they grapple with the improvement program, many sites will move from avoiding, via acting, to adopting, and finally adapting the program.

Box 16.8 **The 4A framework of improvement program adoption**

The 4A framework suggests that subsidiaries, units, or areas *adopt, adapt, act, or avoid* the deployment of a corporate improvement program, as well as its sub-elements and practices:

- "Adopting" means that the unit implements the practices prescribed by the global improvement blueprint in full. When units are new to improvement work, this can be effective.
- "Adapting" happens when the unit alters the practice to achieve a better fit with the local contingencies. This is where you want the units to end up.
- "Acting" happens when the unit pretends to have implemented the practice, but in reality, it is mostly superficial. This is a waste of time and efforts.
- "Avoiding" happens when units avoid the global improvement blueprint or its practices altogether and continue with "business as usual." While this is not an objective, at least it does not incur costs.

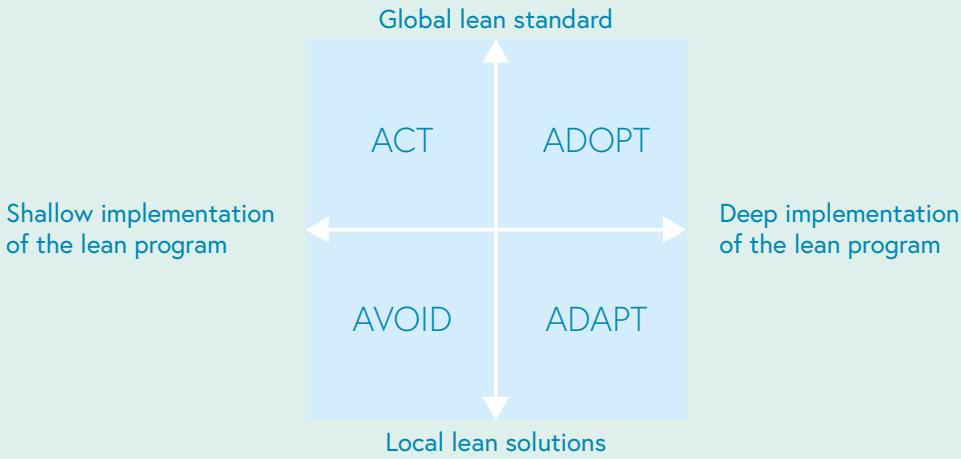


Figure 16.26 The 4A framework for improvement program adoption

51 Netland, T. & Aspelund, A. (2014). Multi-plant improvement programmes: A literature review and research agenda. *International Journal of Operations & Production Management*, *34*(3), 390–418.

16.6.3. The role of a plant improvement department

The evidence is clear that the transformation of a company can only be successful as a result of a strategic decision by top management, with personal involvement by the CEO. It cannot be delegated to a "VP of Lean," "VP of Continuous Improvement," or "VP of Digital Transformation." It must instead be built and sustained by consistent leadership over a long time, so that it becomes part of everyone's regular work. It is not easy but worth the efforts.

Figure 16.27 shows what happens when companies delegate improvement to a department or team of specialists. Not only does the lack of ownership from operations jeopardize the success of every project but the group of specialists cannot possibly be large enough to take on all the necessary work.

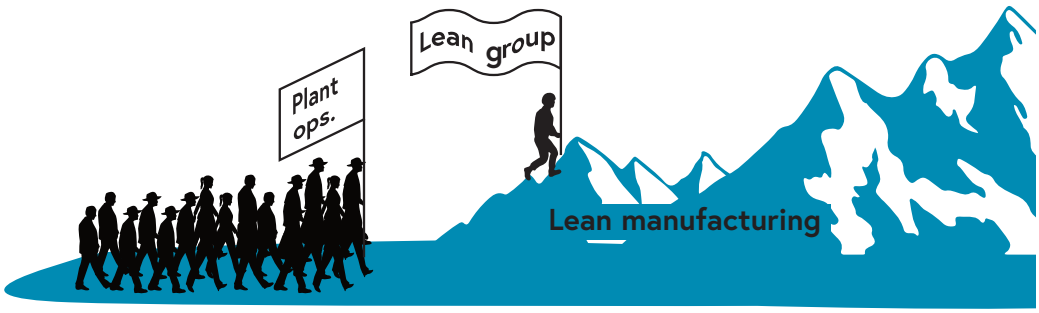


Figure 16.27 What happens when you delegate lean transformation to a department

Most organizations set up a corporate office to select, manage, and execute transformation projects. The projects are done *to* the operations groups they affect instead of *by* them. Kaizen projects must be led by the managers directly in charge of the target process. Not only is this necessary for results to stick but there is no way a corporate department can have enough resources to take on the transformation of a plant. For this to be possible, the line management structure must be large enough for its members to have the time required.

The support department should *support*, not *execute*. It can provide technical assistance and training, make consultants available, intervene to remove roadblocks, facilitate steering committee meetings, organize project presentations, and "own" the documentation, but it should not assume ownership of improvement itself.

> **Stop and Think 16.5**
> Consider the 4A framework, why do you think some local managers decide to pretend (act) to implement a corporate transformation program in their local units, be it lean or "Industry 4.0"?

Summary

The most essential points covered in this chapter were as follows:

1. Performance improvement is key to sustained competitive advantage.
2. Kaizen is an organizational culture making incremental and radical improvement as part of everyday work.
3. Kaizen events is a project-based method for driving improvements, which is better than nothing but not kaizen.
4. Some of the most common project management planning techniques are PERT and CRA, which both start from the deliver date and plan backwards to identify a start date for project activities. Other proposed methods are CCPM and DSM.
5. Agile project management breaks down the project in small increments of testable deliveries that are reviewed with the customer on a frequent basis, for example, weekly or monthly.
6. The Toyota Production System (TPS) enabled Toyota to outperform their competitors on most operational performance areas (better, faster, cheaper). The "TPS house" consists of the two main pillars, just-in-time and jidoka, built on a foundation of operational stability. To strengthen the human side of TPS, Toyota summarized *The Toyota Way* in 2001.
7. Lean production is a label given to TPS in 1988 and has become a dominant business phenomenon in manufacturing. Many companies look to lean for productivity improvement.
8. Company-specific production systems are companies' tailored improvement programs, most often heavily inspired by TPS and lean.
9. Rapid development of new technologies and the advent of a "4th Industrial Revolution" offer new opportunities to improve productivity. Under labels such as "smart factory" or "digital transformation," companies seek improvement from capturing the potential of new technologies.
10. Improvement is a human activity. Without employee engagement and active commitment by leaders, there will not be sustained improvement.

Test yourself

1. How are kaizen events organized? How do they differ from Toyota-style kaizen?
2. What is Toyota's jishuken?
3. What is a Quality Circle?
4. What is an obeya and when do you need one?
5. What is the difference between PERT and CRA? What does DSM add to the project management toolbox?

6 How does agile project management differ from the waterfall method?
7 What are the key differentiators between TPS and mass production?
8 What is meant by the Toyota Way?
9 What are the five lean principles? What was lean in 1990 and what is lean today?
10 What are the benefits of developing a company-specific production system?
11 What is meant by a "smart factory" program?
12 What is the role of management in improvement programs?

Further reading

Anderson, K. (2020). *Learning to lead, leading to learn: Lessons from Toyota leader Isao Yoshino on a lifetime of continuous learning*. Integrand Press.

Chan Kim, W., Renee, M., & John, P. K. (2011). *HBR's 10 must reads: On change management*. Harvard Business Review Press.

Holweg, M. (2007). The genealogy of lean production. *Journal of Operations Management, 25*(2), 420–437.

Kotter, J. P. (2012). *Leading change*. Harvard Business Review Press.

Kotter, J. P., Kotter, J., & Rathgeber, H. (2006). *Our iceberg is melting: Changing and succeeding under any conditions*. Pan Macmillan.

Liker, J. K. (2020). The Toyota way: 14 management principles from the world's greatest manufacturer (2nd ed.). McGraw-Hill Education.

Netland, T. & Powell, D. (2016). *The Routledge companion to lean management*. Taylor & Francis.

Morgan, J. & Liker, J. K. (2020). *The Toyota product development system*. Taylor & Francis.

Ohno, T. (2012). *Taiichi Ohno's workplace management: Special 100th birthday edition*. McGraw-Hill Education.

Repenning, N. P. & Sterman, J. D. (2001). Nobody ever gets credit for fixing problems that never happened: Creating and sustaining process improvement. *California Management Review, 43*(4), 64–88.

Schaffer, R. H. (1988). *The breakthrough strategy*. HarperCollins.

Schonberger, R. J. (1986). *World class manufacturing: The lessons of simplicity applied*. ASQC Quality Press.

Senge, P. M. (2006). *The fifth discipline: The art and practice of the learning organization*. Currency.

Shook, J. (2008). *Managing to learn: Using the A3 Management process to solve problems, gain agreement, mentor and lead*. TPS Enterprise Institute.

Shook, J. & Marchwinski, C. (Eds.) (2014). *Lean lexicon: A graphical glossary for lean thinkers* (5th ed.). Lean Enterprise Institute.

Sobek, D. K. II & Smalley, A. (2008). *Understanding A3 thinking: A critical component of Toyota's PDCA management system*. CRC Press.

Spencer, J. (1998). *Who moved my cheese? An amazing way to deal with change in your work and in your life*. G. P. Putnam's Sons.

Van de Ven, A. (2004). *Organizational change: The Blackwell encyclopedic dictionary of organizational behavior*. Blackwell Publishers.

Womack, J. P., Jones, D. T., & Roos, D. (1990). *The machine that changed the world*. Rawson Associates.

17 Improving Quality

Quality is foundational for manufacturing performance but means different things to different individuals and organizations. Joseph M. Juran, a founding father of modern quality management, argued that the common definition of quality as "conformance to requirements" was wrong, as it ignored customer needs. Juran's view is now the consensus of leading manufacturers. Quality is in the eyes of the customers. As it is neither observable nor controllable in production, we rely on *substitute* characteristics that are variables we can measure or attributes we can observe. Any unit that fails to meet requirements in terms of substitute characteristics is defective but it doesn't mean that every unit that *passes* meets customer needs.

Quality is one of several dimensions of competitiveness and, contrary to long-held beliefs, is not necessarily achieved at the expense of performance in other dimensions. Actually, the ability to produce quality products positively affects flexibility, speed, and productivity. The approach to quality improvement in a plant should be dictated not by a philosophy, but by the quality problems the plant actually faces.

This chapter discusses the management of quality in manufacturing and what can be done to improve it. It starts with outlining an appropriate quality strategy and ends with an introduction to the rich set of the technical methods used to improve quality. Quality has been addressed in many ways, from interchangeable parts technology, to statistical methods and data science, usability engineering, mistake-proofing, and problem-solving frameworks. Like with many issues in manufacturing, there is no one-size solution to quality management.

Learning objectives

After working through this chapter, you will be able to do the following:

1. Explain the role of quality in manufacturing.
2. Give a brief historical account of quality management.
3. Explain the role of process variation for quality outcomes.
4. Use Statistical Process Control (SPC) and understand its limitations.
5. Use Design of Experiments (DOE) to improve a process.
6. Understand the role of data science in the next era of quality improvement.
7. Build-in quality in processes.
8. Use mistake-proofing mechanisms to avoid quality errors.
9. Use Quick Response Quality Control (QRQC).
10. Manage quality mistakes that slip through to the next process and customers.

DOI:10.4324/9781351110310-22

17.1. Introduction to quality management

Into the early 20th century, high-end, luxury products have been described as being of "higher quality" than low-end, basic products. It is no longer used in this sense. The quality of a product is not based on its position within a product line or the market it serves. It is strictly a function of how well it serves the customers who buy it. A luxury product can be deficient in quality and a basic product can excel.

The most commonly accepted definition of product quality is attributed to Joseph M. Juran:[1]

> **Definition**:
> Quality is the ability to meet customer needs.

Juran's view is that quality is *subjective*. He also said that "quality is the agreement of reality with expectancy." The second statement is broader in that it applies not only to manufactured products but to any activity. In everyday language, "expectancy" is the expectation that a product will perform as advertised; in probability theory, it is the expected value of a random variable, as in "life expectancy." Juran's statement works in both senses of the word. A customer buys a product in the hope that it will do what the seller said it would, and it is achieved when all its characteristics match their expected values.

The view of quality as product life expectancy, however, has led to a reductionist interpretation of quality. For example, Philip Crosby defined quality as follows:[2]

> **Definition**:
> Quality is conformance to requirements.

This definition is *objective*. Requirements are formal. They are numerical characteristics of the product that must be within specified tolerances. To Crosby, the manufacturer's responsibility ends with meeting specs, which is much easier to control and manage than customers' often unstated expectations. To Juran, reducing quality to conformance to requirements is simplistic.

> **Stop and Think 17.1**
> You bought a washing machine with a guarantee for 5 years that lasts 6 years before the first part breaks down, but from time to time clothing pieces get stuck in the drum. Is it a high-quality washing machine?

As illustrated in Figure 17.1, *high quality* according to the objective definition may be *poor quality* in the subjective definition. In pastry, the proof of the cake is in the eating, but it is destructive testing and a pastry shop cannot eat all its cakes. To check their quality, it has to use measurable substitute characteristics, also known as proxies. If its diameter is off and it contains half as much sugar as it should, then it is certainly defective. On the other hand, just because it has the right diameter and sugar content does

1 Gryna, F. M. & Juran, J. M. (1988). *Juran's quality control handbook*. McGraw-Hill.

2 Crosby, P. (1979). *Quality is free*. McGraw-Hill.

not mean it tastes good. While conformance to requirements can save you from an expensive lawsuit, the ultimately important criteria is the customers' subjective experience.



Figure 17.1 Objective versus subjective measures of quality

For many manufacturers, their reputation for quality is the crown jewels, and management does whatever it takes to maintain it. To maintain and improve manufacturing quality, they need to understand customer expectations, modify the internal processes, communicate their own expectations to suppliers, and help suppliers meet them.

In some markets, customers have low expectations. It can be because protectionism has shielded them from knowledge of world-class quality or because the products are innovative and the customers don't know what is possible. In such markets, companies succeed by flooding the market with goods first, and worrying about quality later.

17.1.1. About defects, defectives, and errors

Terms like defect, defective, and error occur all the time in discussion of quality, and we must be clear about what we mean by them:

- A **defect** is a characteristic a product unit should not have.
- A **defective** is a unit of product unit with at least one defect.
- An **error** is committed by a human accidentally doing the wrong thing.

Errors cause *defects* but not all *defects* are the result of *errors*. In an interactive teaching tool, W. Edwards Deming demonstrated this point in his famous "red bead" experiment. Even though "willing workers" want to do a good job, their success is dependent on system design.

The Japanese literature, both in translation and in the original, frequently confuses defects and defectives. The distinction, however, is sharp. For example, if a transmission case with 3.4 defects per million opportunities (dpmo) has 2,200 critical dimensions, and therefore 2,200 defect opportunities, it can amount to 7,452 defective parts per million.

17.1.2. **A brief history of quality**

Cicero coined the word "qualitas" in his book *Academica* in 45 BCE, as a translation of the Greek "poiotetes." What Cicero meant is not obvious but he seems to have been a systems thinker, in that his version of quality was what makes a system more than the collection of its parts. Over the ensuing 2,000 years, quality has taken on a variety of meanings, mostly not related to manufacturing.

In manufacturing, the story of quality is a centuries-long struggle for consistency of output, from when instructions told blacksmiths to heat the work piece until it is "cherry-red," to the current ubiquitous sensors connected to the Internet of Things (IoT). As a rough cut, we see the following eras in the evolution of quality methods and practices:

1. Interchangeable parts technology (1760–1920)
   - Engineering drawings
   - Machine tools
   - Critical dimensions and gauges
   - Tolerances and allowances
2. Statistical quality control (1920–1945)
   - Process capability and Statistical Process Control (SPC)
   - Inspection and acceptance sampling
   - Quality Control and Quality Assurance
3. Post-war developments (1945–1980)
   - Total Quality Control (TQC)
   - Application of Statistical Design of Experiments (DOE) to manufacturing quality
   - De-emphasis of quality in the US, leading to stagnation and decay
4. Recent developments (1980–2020)
   - Total Quality Management (TQM)
   - Development of ISO standards
   - Six sigma
   - One-piece flow and rapid response
   - Mistake-proofing
   - Usability engineering
5. Current developments (2020 onwards)
   - Data science

Interchangeable parts technology (1760–1920)

The earliest efforts at improving manufacturing quality as we know it went by other names. The development of *interchangeable parts technology* in 18th century France and 19th century US *was* a quality improvement effort, targeted at making goods consistent enough to eliminate custom adjustments whether in original assembly or in part replacement. Its legacy includes the concepts of engineering drawings, critical dimensions, tolerances, and gauges that are the foundation of quality management, as developed since the 1920s. It is ignored in the recent literature on quality but Walter Shewhart,[3] the inventor of statistical quality control, acknowledged it explicitly.

3 Shewhart, W. A. (1939). *Statistical method from the viewpoint of quality control*. Dover Publications.

Statistical quality control (1920–1945)

On the foundation of interchangeable parts, the next challenge was to establish and maintain *process capability* – that is, the ability to produce parts that are consistently within specs. Working in the high technology of the 1920s, Shewhart at Western Electric applied state-of-the-art statistics to develop control charts that generate alarms when processes drift, preferably before the drifts are high enough to make parts defective. By the end of World War II, W. Edwards Deming, Joseph M. Juran, Abraham Wald, W. J. Jennett, and others, had rounded out the body of knowledge that is, to this day, taught as *Statistical Process Control* (SPC).

Concurrently to SPC, Harold F. Dodge and Harry G. Romig developed *acceptance sampling*, to determine the quality of a batch of products by selecting a specified number for testing or inspection. Acceptance sampling was particularly helpful when a test required the destruction of a product, when the cost of inspection is very high, or when the inspection takes a long time. Acceptance sampling was first applied in the US Military to bullets during World War II.

Post-war developments (1945–1980)

During the previous era, the development of quality management was led by the US. World War II had been a period of rapid innovation, with, for example, TWI introduced to train a new workforce of highly motivated women, a moving assembly line for bombers in Willow Run, Liberty Ships built by arc-welding vertical hull slices, and the invention of new statistical methods in quality. But after World War II, while quality was less emphasized in America it became the main focus of Japanese industry. The American occupation of Japan after the war included the transfer of quality knowledge to help rebuild Japanese industry. Japanese manufacturers not only implemented the quality control techniques taught by Deming, Juran, and others, but also enhanced them.

Post-war developments included Armand Feigenbaum's *Total Quality Control* (TQC)[4] in 1951, in which "Total" meant "affecting all functions." TQC is now better known for its Japanese version formulated by Kaoru Ishikawa[5] in 1968, in which "Total" means "with participation by every employee." The Japanese version of TQC is also called *Company-Wide Quality Control* (CWQC).

In Japan, the *Deming Prize* has been awarded since 1951 to companies and individuals for contributions to quality. In 1951, Juran also compiled the first *Quality Control Handbook,*[6] whose 7th edition came out in 2016, and formulated the approach now known as the Juran trilogy, of quality planning, control, and improvement. Other contributors, like Dorian Shainin and Genichi Taguchi, adapted statistical *Design of Experiments* (DOE) to the solution of quality problems.

By the late 1970s, quality in American manufacturing had slipped from an overwhelming advantage to a handicap in cars, consumer electronics, shipbuilding, or steel. In 1980, NBC broadcast a documentary called *If Japan Can, Why Can't We?*,[7] introducing W. Edwards Deming to the American public. It was a wake-up call, discussed for years afterwards around cafeteria tables in American factories.

4 Feigenbaum, A. V. (2008). *Total Quality Control.* I.K. International Publishing House.

5 Ishikawa, K. (1985). *What is total quality control? The Japanese way*. Prentice-Hall.

6 Juran, J. M. & Gryna, F. M. (Eds.) (1988). *Juran's quality handbook: The complete guide to performance excellence* (4th ed.). McGraw-Hill.

7 NBC (1980). *If Japan can, why can't we?* URL: https://www.youtube.com/watch?v=vcG_Pmt_Ny4

Recent developments (1980–2020)

In the early 1980s, quality became once again a focus of US and European manufacturing. Progress was initially slow. In automotive, until the 2010s, American and European brands were lagging behind Toyota, Lexus, and Honda in the Initial Quality Surveys (IQS) conducted by J.D. Power and Associates. By 2020, however, American, European, and Korean car makers had caught up with Toyota, Honda, and other Japanese auto makers.

In the 1980s, the US Department of Defense started mandating "*Total Quality Management*" (TQM) in its own operations, and this term gradually replaced TQC in the literature. It has since taken on different meanings for different organizations:

- In Japan, TQM is mostly associated with the theories of Noriaki Kano, developed in the 1980s and taught by JUSE. It is also known as the *Kano Model*, which is a 2x2 roadmap that help prioritize what product features to develop based on what will delight customers.
- In the US, TQM is primarily understood to be what the *Malcolm Baldridge National Quality Award*, established in 1987, is given for. Quality appears in only one of the seven criteria used to grant the award.
- In Europe, TQM was rebranded and formalized to the *European Foundation for Quality Management* (EFQM) in 1989.
- Worldwide, TQM is associated with the *ISO-9oox* family of standards, certification in which has become a cost of doing business for many manufacturers. Evidence that it has contributed in any way to manufacturing quality is hard to find. Its most lasting legacy may be that it has forced companies to pay more attention to master data management.

Later the quality movement extended to the whole organization when companies started to develop *Integrated Quality Management Systems*. Today, quality initiatives are often subsumed under broader improvement programs, such as for example corporate lean programs, *lean six sigma*, or *company-specific production systems*. When thinking of manufacturing as a system, it is clear that there is synergy between quality and other competitive capabilities like speed, sustainability, and flexibility.

Meanwhile, in Japan, quality improvement in mature industries took a different direction, away from sampling and from what Shigeo Shingo[8] called "inferential statistics," and towards 100% checking of parts through go/no-go gauges, successive inspection, and mistake-proofing (poka-yoke). *Within their range of applicability*, these methods achieved levels of quality that were far superior to the statistical methods.

Current developments (2020 onwards)

In high-technology processes the competitive achievement of process capability requires the state of the art in *data science* to identify and capture hidden relationships in manufacturing data, not only when they are numerical and structured but also when they come in the form of unstructured text, pictures, or videos.

8 Shingo, S. (1986). *Zero quality control: Source inspection and the poka-yoke system*. Productivity Press.

A place to start is visual inspection systems powered with *artificial intelligence* (AI). Today, such systems can be implemented at local stations and almost out-of-the-box. Supervised AI systems use annotated photos of quality errors on products and learn how to recognize them. Unsupervised AI systems learn from experience. Visual inspection systems can be coupled with automation and replace manual visual inspection, but it has not done so in all cases, perhaps for two reasons:

1. AI does still not have the intelligence and agility of a human. Some errors such as wrinkles on textiles or sealing contamination are notoriously difficult for AI to pick up.
2. Even when the systems works, we would like to have a human there to handle the rejected parts, do root cause analysis, and prevent the error from happening again.

Data mining is the analysis of data that was collected for a different purpose, as opposed to Design of Experiments (DOE), where the data is collected specifically for the purpose of answering a question. The term may be recent but the activity has a long history. In manufacturing, it can be used today to find root causes of defects with AI[9] or drawing process maps of actual product or information routing as in *Process Mining*.[10]

17.1.3. The right quality strategy

The approach to quality improvement in a factory should be dictated by the quality problems it is facing, not the blind copying of a quality improvement method. Box 17.1 and Box 17.2 illustrate two radically different examples: One concerning high technology in the semiconductor industry and one low-technology process in the automotive industry. When the main quality problem is operator errors, then their prevention calls for mistake proofing. But if the main problem is the inability to hold tolerances consistently, spending your resources on mistake-proofing would be lining up deck chairs while the ship is sinking. Conversely, data science is not much help in the prevention or errors that occur once or twice a year.

The right approach to quality management is tailored to specific needs, but there is a general sequence to how problems should be addressed:

1. **Process capability.** If your main problem is lack of process capability, your failure rate will typically be high, and this issue should be addressed ahead of anything else. Beyond this point, it's no longer the identification of root causes that matters most but instead the speedy detection of problems.
2. **Problem detection time.** The next order-of-magnitude improvement in quality performance comes from moving to one-piece flow with first-in-first-out (FIFO) sequencing. The proportion of defectives traced to the operations in a cell drops by at least 50%.
3. **Human error.** The effect of one-piece flow and FIFO in turn levels off at percentages of defectives commonly on the order of 1% to 0.1% = 1,000 parts per million (ppm) defects. It is not by itself able to achieve low ppm rates. Human error still affects individual parts, and this is where mistake-proofing comes in: It is the strategy that takes you from, say, 1,000 ppm to 15 ppm.

9 Lee, J. (2020). *Industrial AI: Applications with sustainable performance*. Springer.

10 van der Aalst, W. (2016). *Process mining: Data science in action*. Springer.

Box 17.1 **A diffusion furnace in semiconductor wafer processing**

In semiconductor chips fabrication (Figure 17.2), a diffusion furnace is used to add doping impurities into silicon wafers. One load of 200 wafers is typically worth about $250,000. This is an industry in which products are obsolete before processes reach the maturity that is the starting point of high-volume production elsewhere.

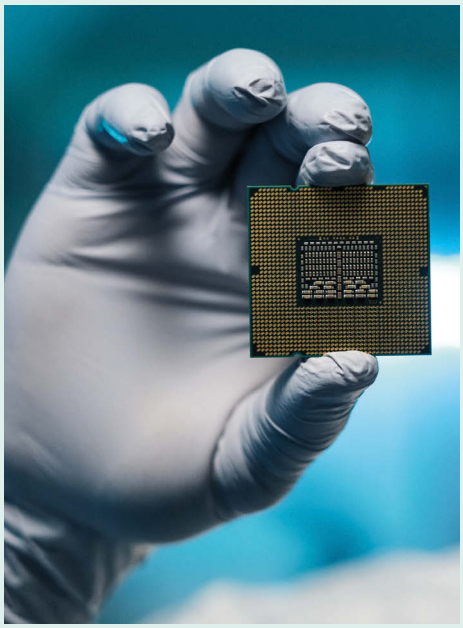


Figure 17.2 A semiconductor chip

These machines typically have difficulty producing consistent doping profiles from one end of the batch to the other, and engineers keep trying to extend the "flat zone" in the middle where the results are within specs. In addition, small fluctuations in the results of this operation interact with process variables set at upstream and downstream operations to cause circuit malfunctions that may not be detected until final test.

These issues are addressed by process and yield enhancement engineers. The process engineers who run experiments understand solid-state diffusion and thin film deposition, as well as statistical Design of Experiments (DOE), and data science, from multivariate analysis to deep learning. What they do clearly has to do with controlling the characteristics of the outgoing product and eliminating the production of defectives. It is an essential function in this business, and it is centered on data science.

This, however, does not exclude the presence of human error. The selection of the process program can, for example, be based on an operator reading a run sheet and typing in an identification code. For example, in one factory, an operator loaded the wrong process program four times a year, causing an entire load to be destroyed, and costing $1M/year. How could this be solved?

In semiconductor manufacturing, you find engineers specialized in computer systems to control process equipment. The embedded control systems are themselves PCs on most machines. They communicate with outside computers using industry-specific protocols. Using commercially available packages, in-house computer engineers can program an external PC to communicate with the embedded controller on one side and with a plant-wide system on the other to prevent the wrong process program being loaded and executed.

A $50K external controller was bought. It was an expensive mistake-proofing device, but it paid for itself immediately. It may be expensive, but it is cost-justified and implementing it does not take resources away from process engineering and yield enhancement. In this case, mistake-proofing and data science are complementary and both relevant, because quality losses occur both for lack of process capability and for human error.

Box 17.2 **Oil filling of automobile parts**

Figure 17.3 shows automotive parts, weighing about 9 kg (20 lbs) each, that must be filled with 4.5 dl (1 lb) of oil prior to assembly. This operation doesn't require an expensive machine. At most, we may have one engineer looking after this operation some of the time.



Figure 17.3 Batch method of oil filling

If we have process capability issues to contend with, we address them first. Once it is under control, human error becomes the next focus of improvement. If we know how to dispense a pound of oil accurately and consistently, then the parts coming out of this operation can be defective only if the operator omitted to fill it or filled it more than once.

Figure 17.3 shows a filling method that is conducive to this kind of error. The operator works on a batch of parts arrayed on top of a cart, locates the dispensing nozzle in the filling hole and presses a trigger. The controller then dispenses the specified amount and beeps when done, which causes the operator to index over to the next part. The work position is ergonomically questionable, and filled and empty parts coexist side by side. The slightest distraction can cause the operator to forget where she is, and the difference in weight between an empty part and a filled part cannot be felt by lifting it.

Contrast this with the filling station shown in Figure 17.4. Setting up one-piece flow in front of a fixed dispenser eliminates most error opportunities. It cleanly separates the unfilled units from the filled ones and makes sure every unit gets a turn on the filler.

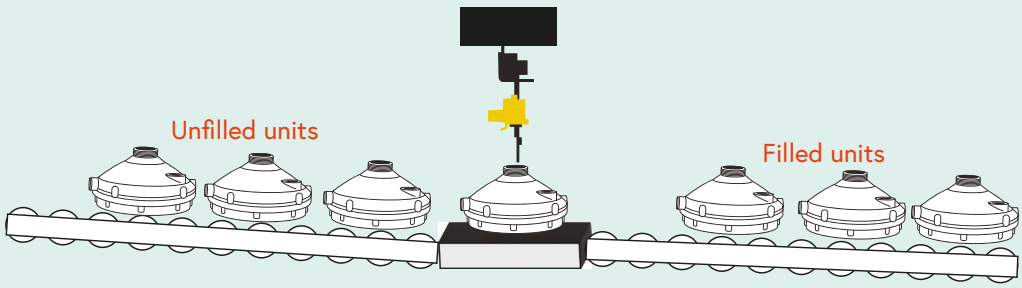


Figure 17.4 Sequential filling station

This eliminates most but not all the errors, since the operator can still move an unfilled unit forward. A person cannot feel the weight difference, but an automatic check weigher can. So in a second round of improvement, a check weigher can be retrofitted to the dispenser to disallow double filling and prevent the unloading of unfilled parts. This mistake-proofing device then eliminates the last opportunities for errors.

This is a case where statistical methods make no contribution to the solution. At best, they can be used to assess the frequency of errors.

Table 17.1 summarizes this rule-of-thumb. The figures in the left column are to be interpreted as orders of magnitude. They correspond to figures observed in automotive parts companies. Rapid product obsolescence in high-technology manufacturing companies may prevent them from ever experiencing the stages described beyond the first row.

Table 17.1 Quality levels and improvement approaches

Defective parts	Quality problem	Solution
30% to 3%	Process capability	Process analysis via data science, process engineering, DOE
3% to 0.1%	Problem detection time	Bulit-in-quality via flow (assembly cells/lines), FIFO, visual controls, QRQC
0.1% to 15 ppm	Human error	Elimination of quality errors via mistake-proofing

Step 1: Ensure process capability

Low yields and high failure rates are a constant of life in high technology manufacturing. Just as SPC came out of Western Electric, the high technology of the 1920s, six sigma came out of Motorola in a similar situation in the 1980s. This does not mean that the relevance of the approach is limited to high technology. Even in mature industries, there are processes, such as painting car bodies, with double-digit rework rates even among the best manufacturers. Finally, processes temporarily exhibit this type of behavior during new product introduction.

By "statistical methods," most quality professionals understand the combination of control charts, histograms, scatter plots, sampling methods, inspection plans, and other tools collectively known as "SPC" or "SQC." These methods, pioneered 100 years ago by Walter Shewhart and later promoted and enhanced by W. E. Deming, A. Wald, and K. Ishikawa, were targeted at the processes and based on the information technology of the era. The production of these charts continues today as a result of customer mandates enforced by audits, but they play almost no role in problem-solving, even where statistical methods are needed.

Since the 1930s, not only has data acquisition technology changed, but statisticians have developed many tools that take advantage of computers. Semiconductor testers gather hundreds of variables on each circuit, but, even in mature industries, spindle probes in machining centers, Coordinate Measuring Machines (CMMs), or Supervisory Control and Data Acquisition (SCADA) systems can collect numerous product characteristics and feed them immediately to analysis software.

Even the electronic spreadsheets found on every engineer's desk come with a data analysis package that is far more powerful than traditional SPC. Beyond this, specialized statistical software products supporting multivariate exploratory data analysis and Design of Experiments are available for less than the cost of one engineer-week. Yet, except in special industries like semiconductors, the power of these tools remains largely untapped. The reason is not technology but lack of skills. Not only are people who understand tools like the Fisher discriminant or fractional factorial designs of experiments not commonly found in factories, but their knowledge is of limited value unless they also understand plasma

etching, the appropriate feeds and speeds for milling titanium, or whatever other process they are using. And people combining data science *and* process knowledge are almost nowhere to be found.

Step 2: Built-in quality

Once process capability is established, the plant faces a change in the nature of quality problems, from drifts in process parameters to infrequently occurring *discrete events* like tool breakage, skipped steps, or mislabeled components. The challenge at this point is to detect them quickly and react promptly. The problems that require an elaborate root cause analysis, by definition, have been solved already, and now speed takes precedence over analytical sophistication.

The solution is to *detect quality problems faster* by making them appear immediately or shortly after the process. The way to do it is to seek to implement one-piece flow, first-in-first-out (FIFO), visual control, and in-line quality control through manual or computerized visual inspection or go/no-go gauges. The goal is to never pass on a mistake to the next process but these methods do not guarantee it doesn't happen. Therefore there is also the need for problem-solving techniques that handle deviations quickly. One such method is Quick Response Quality Control (QRQC).

These methods are ineffective on processes with double-digit percentages of defectives because they would make the system break down. High volume production with an unstable process *requires* buffers between every two operations, simply to protect the downstream operation against output fluctuations upstream. The process characterization and stabilization work done with data science enables one-piece flow and FIFO, which in turn enables the process to reach the next plateau.

Step 3: Mistake-proof processes

Even if all the above is implemented, humans still make mistakes. The only way to avoid it is to make it impossible to err. *Poka-yoke* designates small equipment retrofits that mistake-proof production processes. Where most defects are caused by human error, poka-yoke can get your reject rate from 0.5% to <1 ppm. Human error is particularly notorious in mechanical and electronics assembly operations.

Step 4: Manage rework and recalls

We marvel at the ingenuity of poka-yoke devices, but we have yet to invent one for every defect opportunity, and, until we do, must rely on less powerful tools. The inevitable result is that poor quality sometimes slips through to the next process or, worse, to the customer. How we handle quality mistakes deserves a strategy on its own,

For *rework*, we must decide where to do it, in the line or at the end of the line, set the threshold between scrap and parts to rework, and how to manage them. Every rework is waste, but sometimes it is even done on purpose. For example, during the global chip shortage in 2021, several vehicle manufacturers produced semi-finished vehicles to stock, with the idea to rework them when the needed chips got available again. While this is something that should be avoided, when it first happens, one needs to balance the best use of available resources. In this case, sending the workforce on paid leave was considered an even greater waste and a more risky option than to build semi-finished products.

Defectives eventually and almost inevitably slip through to customers. If the defects are systematic and safety-critical, manufacturers have to issue a *recall*. Recalls are usual in pharmaceuticals and food and beverage industries, but also in high-technology industries like healthcare technologies, automotive, and aerospace. It is also not unusual for retail chains to recall products from low-quality suppliers of all kinds of consumer goods such as toys, white goods, toasters, etc. Managing recalls well can turn a negative customer experience to a positive experience.

17.1.4. **Organizing for quality**

However, even in organizations where all members are responsible for the quality of their output, you still need a Quality department to do the following:

1. Interact on quality matters with customers and suppliers.
2. Organize and lead the organization's response to quality emergencies.
3. Monitor quality practices inside the organization and raise alarms when they slip.
4. Calibrating measurement instruments.
5. Secure and maintain required certifications.
6. Watch for new developments in technical tools and methods of quality management.

Compared to the Quality Assurance (QA) department of a mass production organization, with its emphasis on inspections done by QA technicians, a Quality department with the above charter has about half the size and twice the authority.

The Quality department

Customers issue quality problem reports to the company, but not to the manager in charge of a shop floor operation. These reports land in the Quality department, which routes them where needed and reports back to the customer on countermeasures and permanent solutions in the format mandated by the customer. The shoe is on the other foot with supplier quality problems.

In emergencies, the quality manager has the authority to stop production and organize recalls. When a firm responds to a quality emergency, external and internal quality have to be managed jointly, because the firm simultaneously has to interact with aggrieved customers, stop further shipments of defectives, and solve the problems that caused them.

In some companies, during product launches, the Quality department sends technicians to dealership to observe customer returns, and customers' reactions to the product's features, which results in detecting and solving problems early.

In response to customer mandates, the Quality department ensures that the company is properly certified. Some customers will mandate that suppliers must be ISO certified. It doesn't improve quality per se, but it has become a cost of doing business. Someone has to be in charge of securing and keeping that certification, and that job falls to the Quality department, which minimizes the impact on everyone else.

Management changes and employee turnover always put the company's quality practices at risk of slipping, and it is part of the Quality department's job to keep it from happening, by making sure that

up-to-date instructions are posted, that instruments and gauges are properly calibrated, and that mistake-proofing devices are not cast aside.

Finally, quality practices evolve. New technology appears, be it auto-ID or analytics; new management methods appear like Toyota's JKK (jikotei kanketsu, 自工程完結, or "autonomous process completion") or Nissan's Quick Response Quality Control (QRQC). It's the Quality department's job to learn and recommend the implementation of relevant innovations in the company.

Who is responsible for quality?

The problem is not the existence of the Quality department but the tasks it has been traditionally assigned, essentially relieving production of responsibility for quality. You could produce defectives and rely on inspectors from the Quality department to filter them out. Then, when defectives escaped to customers, it was the inspectors' responsibility. It used to be a widely accepted practice. We now know that it is more effective to make everyone responsible for the quality of their own output. Even when you use a machine or assemble purchased components, you are supposed to take action as soon as you notice defectives, for example by stopping the line.

In society at large, it is fair to say that fire prevention is everybody's responsibility but it doesn't mean you do without professional firefighters. No matter how dedicated everyone in the company is to quality, you will still have instances like horse meat mixed with beef in processed foods, unintended acceleration in cars, or rockets that blow up. The question is, "Who is the first responder?" A problem report arrives from the outside world; "Who receives it?" There has to be a department within the company to take these calls, take immediate countermeasures, organize the search for a permanent solution, and communicate with customers.

There is plenty of other work to do to achieve and maintain quality. You start a new fabrication process with, say, 30% of defectives. You bring it down to 3%, then 0.3%, then 500 ppm, then 20 ppm, then 1 ppm, and so on. Whatever quality level you achieve, you must maintain it as new people join the production team while machines and tools age. Meanwhile, the external mandates don't end, and then you have to start over with the next generation of products.

The Quality department will exist as long as there is work to be done in that area that requires a special set of skills and is best done centrally, as opposed to delegated to all other departments.

Quality and public relations

Another responsibility with respect to the outside is nurturing the company's reputation for quality. In many markets, it is the crown jewels. Delivering high quality is not enough; excellent performance also has to be publicized, and the company's reputation sometimes needs to be defended against attempts by competitors or counterfeiters to ruin or take advantage of it.

There have been cases of competitors making and selling defective copies of a company's products in order to ruin its reputation. In the 1910s, NCR did it to corner the market for cash registers, for which future IBM founder Thomas J. Watson was convicted of illegal anti-competitive sales practices. In the 1970s, a supplier of taxi meters, Safaa, was driven out of the Egyptian market when taxi drivers started

reporting malfunctions, which in fact were in counterfeit meters from a competitor and designed to ruin Safaa's reputation. Today, social media are an arena where a company can receive not only genuine feedback on quality, but also be the target of smear campaigns.

You can move towards an ideal state in your *internal operations*, the part of the world that is under your control, but it's not enough to excel in it. The company and its products must be known in the market to excel. The external environment is not under your control, and reputation management is part of quality management.

Internal versus external responsibilities

Managing quality requires many actions that involve *external* interactions, like monitoring product launches, organizing recalls, or combatting smear campaigns. It also involves multiple constituencies, including not only customers but also suppliers, complementors, competitors, communication media, and governments.

The need to worry about supplier quality is obvious but the Purchasing department is not properly equipped to vet potential suppliers for their quality or help existing suppliers improve theirs. It needs assistance from the Quality department.

Complementors are companies that sell to your customers products that enhance the value of yours. If you sell computers, vendors of software that runs on your computers are complementors. These are companies you make alliances with. They have a stake in the quality of your output, as you do in theirs. When a software product crashes hardware, it blemishes the reputations of both the software and the hardware vendors.

Competitors do not always have an interest in hurting your reputation for quality, for example when there is an overriding need for the public's trust in the entire industry. Airlines do not raise doubts about the safety of other airlines, because they would suffer too if the public lost confidence in aviation. In this case, you need to organize some level of collaboration with your competitors through industry groups or professional societies. And government may step in with regulations if quality problems with your products put the public at risk. For all these reasons, most organizations find it necessary to have a Quality department.

> **Stop and Think 17.2**
> Are all quality-related activities improving quality? Why or why not?

17.2. Inspection and testing

Regardless of how well you build quality into your products, inspection and testing activities do not vanish. They may be reduced in scope but there are few assembled products that are shipped without any testing, and there are few manufacturing processes that do not require some form of inspection somewhere. The lean manufacturing literature may convey the impression that inspections, testing, and

rework have been completely eliminated. In electronics, testing often takes longer and uses more expensive equipment than assembly. It is not unusual to have a testing process after each assembly step. Even Toyota feels the need to organize a systematic effort at finding defects in incoming raw materials, WIP, and in finished goods.

While serving the same purpose, inspection and tests differ in content. Inspecting a part or product unit means observing or measuring some of its physical characteristics, while testing is focused on how it functions. Inspecting a mechanical assembly may mean verifying that all the screws are present and properly torqued, and that a bead of sealant of the right width is present where a head meets a housing. A test on the same part might entail pressurizing it under water to check for leaks. Tests are usually more directly related to function and involve less subjective assessment. Automatic testing is much more common than automatic inspection.

Inspection and test operations are usually part of assembly processes, because a substantial part of the value of this process to the customer is that the supplier takes responsibility for the outcome. While some describe these activities as "non-value added," customers disagree:

- In aerospace and defense applications, the government demands inspection and test reports to be shipped as part of the product.
- In consumer goods, companies use the thoroughness of their inspections as an advertising argument.
- In computer assembly, testing and "burn-in" are the key reason there are so few do-it-yourselfers.

17.2.1. The rational and problems of inspection and testing

In general, inspections serve several purposes:

1. **Filtering.** The farther a defective part or product is allowed to proceed, the more expensive it is to deal with. The inspection or test serves to sort the good from the bad to keep from shipping defectives to the next operation or to a customer, or to avoid accepting defective parts from a supplier.
2. **Binning.** An inspection or test can result in placing each unit in an appropriate bin with each bin corresponding to slightly different customer specs. This is common in electronics and has both a technical and a commercial dimension. A part can meet several overlapping sets of specs, and the company makes a business decision as to which bins it actually goes into.
3. **Troubleshooting.** Changes in the rate or mix of defects are warning signs of changes in the process itself. Sometimes, the primary purpose of inspections and tests is process diagnosis for yield enhancement. This requires retaining inspection results or test data and mining this data for information about the process, particularly about non-obvious interactions between different operations.
4. **Learning.** Even on lines that are running well, units of finished goods are periodically pulled and subjected to extensive and possibly even destructive series of tests. During product development and pilot production, learning is actually the primary purpose of inspection and test. In this case, the data produced may be considerably more valuable than the parts. Data collection and database structures must be specifically designed for this purpose
5. **Complying with external mandates.** When an inspection is conducted for the purpose of satisfying external audit requirements, then the key issue is that the procedure must match the mandated standard.

From the point of view of shop floor operations, inspection and test activities are disruptive in multiple ways:

- **Inspection or test operations vary in duration.** Since a defective unit is rejected as soon as a defect is found, its processing takes less time than a good unit that undergoes all tests. As a result, capacity planning and sequencing are more challenging for these operations than for normal production tasks.
- **They generate a flow of rejects that must be managed.** Rejected units may be passed onwards after further testing, reworked, or scrapped. Technical resources are needed to make the decisions and carry out the needed rework, and logistics resources are needed for final disposition.
- **They generate information about the manufacturing process.** Inspection and tests are a major source of information about process conditions. It must be recorded, communicated to engineering, and acted upon promptly. This requires an information system tailored to handle not just part quantities but technical data as well.
- **Their psychology is different from that of production.** Finding fault in what has been made requires a different mindset than making it.

17.2.2. Quality inspection of incoming materials

Not all suppliers are equal. Some are outstanding performers in quality; others, on notice that they will be terminated unless they improve. Different policies are needed for different suppliers. The best suppliers probably don't need any incoming inspection, and the whole receiving procedure may be limited to scanning the barcode of the shipment. They may deliver 10 times a day, in quantities that go straight to assembly.

At the opposite end of the spectrum are problem suppliers. As long as they are used, Receiving needs to check everything they do. Is it the right item? Is it in the right quantity? Do all units work and meet specs? No sampling inspection will do. Every part must be checked. And the incoming materials warehouse can dedicate a zone to this supplier, to easily monitor all its items. Figure 17.5 illustrates the labor-intensive process. Then, there are levels in between.

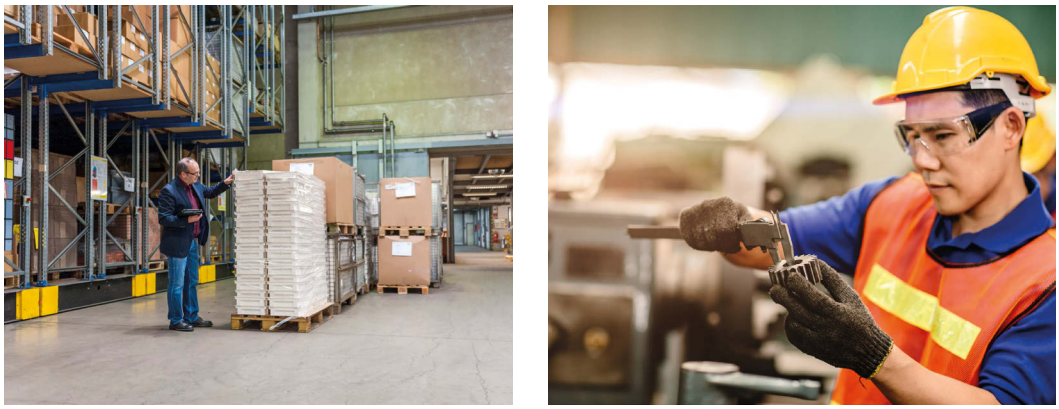


Figure 17.5 Inspection of quantity and quality of parts from suppliers

If quality issues alone drove the sizing of deliveries, they would be in single units, at intervals matching the takt time of the products they go into. From a strict quality perspective, it would minimize the customer's risk of holding defective parts or parts made obsolete by an engineering change order, and whenever a supplier would deliver a defective part, feedback about it would arrive before many more are made. While logistical constraints force customers to receive lots of more than one unit, milk runs can keep lot sizes down, and having different policies depending on supplier performance enables a

manufacturer to contain the cost of receiving small lots, since the bad performers requiring attention are few, and either on the mend or on the way out.

The customer's quality assurance group needs to think through what it can or should require from suppliers, given that it cannot possibly have a deep understanding of the issues faced by hundreds of suppliers working with metals, plastics, glass, ceramics, electronics, etc. Outgoing quality performance can be required, as can traceability and responsiveness in case of emergencies but it is often counterproductive to prescribe *how* suppliers should go about meeting these requirements. The customer organization has learned tools and techniques that have worked – and not worked – in solving its own quality problems. Sharing this experience with suppliers is providing genuine support, but the managers and engineers in supplier organization should be treated as professionals making their own decisions on which methods to use for what purpose in their own plants, and be judged exclusively on the results.

17.2.3. **Acceptance sampling**

In acceptance sampling, manufacturers choose a sample size, design the sampling procedure for representativeness, and fit distributions for the characteristics of interest for the whole batch.

> **Definition**:
> Acceptance sampling is a quality control procedure where a small sample is drawn and inspected in order to make the decision to accept or reject the full batch.

Acceptance sampling is of limited value when defectives are on the order of a few parts per million, as has been achieved by some auto parts suppliers. But it remains a useful tool in a number of other circumstances. As much as we would like to avoid it altogether, there are situations where some form of sampling is inevitable. Even when checking, measuring, inspecting, or testing every workpiece, it is often impossible to cover all possible uses:

- The testing many be destructive, as in the case of eating food products.
- Testing a microprocessor's ability to run every possible program is impossible because their number is infinite.
- Raw materials that are mined out of the ground, recycled waste, or agricultural crops, arrive in batches of variable composition that needs to be assessed from samples.

In *Out of the Crisis*,[11] W. Edwards Deming denounced acceptance sampling as a waste of money and blamed the approach for guaranteeing that "some customers will get defective products." He suggests that one either does 0 inspection or 100%, and proposed the following criterion for a tipping point from 0 to 100% inspection:

$$Tipping\ point = \frac{Cost\ of\ inspecting\ one\ piece}{Cost\ of\ accepting\ a\ defective}$$

For example, if the cost of inspecting one piece is $100 and the cost of accepting a defective is estimated to $10,000, the tipping point is 1/100. If the fraction defective of the incoming units is below the

11 Deming, W. E. (1986). *Out of the crisis: Quality, productivity and competitive position*. Cambridge University Press.

tipping point, for example, every 1 in 200, then 100% inspection doesn't pay off; otherwise, it does. This formula is simple and logical, except that the cost of accepting a defective is not easy to assess. It is dominated by the damage done to the company's reputation for quality, which is not easily reduced to a number.

An intermediate policy is to check the first and last piece of a delivered batch, but it is only feasible if the lot you receive presents the parts in the order in which they were made because, otherwise, you cannot interpolate the characteristics of the parts in-between. It is a method you can apply in your own processes where you practice one-piece flow but rarely on purchased items.

In reality, we cannot accurately calculate the tipping point, but we know that the further downstream defectives are detected, the more damage they do, even if we don't always know how to put a number on it. If defectives pass through Receiving and are detected at Assembly, they disrupt the assembly process, particularly if it involves kitting; if they are detected at Final Test for products they go into, the product units are diverted to a repair area where they may have to be taken apart; if they escape to the market, they may harm customers.

So the real question is whether there is anything we can do, short of inspecting every unit, that can be effective at protecting the downstream process. Acceptance sampling can be one such strategy, but it should never be a company's only strategy to ensure quality.

17.2.4. **Test sequencing**

An inspection or test operation is usually a series of steps, involving dimensional measurements and tolerances or control limits, visually examining the smoothness of surfaces, or making a device perform some of its intended functions. When this is done in production, the operation is stopped on a unit as soon as a defect is found, and only defect-free parts undergo the full sequence. It is possible to sequence the steps to minimize the average total time required to inspect or test a part.

Sequencing by merit

There is a well-known theory on how to do this that is applicable when two conditions are satisfied:

1. **The tests or inspection steps are independent.** Paint scratches and missing screws, for example, are independent, but, on the other hand, the tests of whether a microprocessor runs at 25MHz and at 50MHz are clearly not, since, if a unit fails at the low clock speed, it certainly won't work at the high one.
2. **The outcome is strictly pass or fail, as opposed to binning.** A unit with a missing screw clearly failed, but a microprocessor that won't work at 50MHz may work at 25MHz and be saleable in the corresponding bin.

When conditions 1 and 2 are satisfied, sorting the steps by decreasing values of a "figure of merit" gives you an optimal sequence. The figure of merit is defined in one of the following ways:

$$\text{Figure of merit (step)} = \frac{\%\ \text{rejected}}{\text{Time required}}$$

or

$$\text{Figure of merit (step)} = \frac{\%\ \text{rejected}}{\$\ \text{spent}}$$

Sequencing by decreasing values of the first figure of merit minimizes the time required to reach a decision; by the second figure of merit, the cost of reaching that decision.

Sequencing by induction

When tests are related, then each step gives some information as to what the outcome of the other steps will be. This has led to the use of information theory to generate decision trees. The first example dates back to 1972, but, to this day, this technique is not widely used in manufacturing. The problem in 1972 was to improve the efficiency of kidney disease diagnosis in a hospital using 27 different tests. The science was all known: For each of the 13 diseases, the doctors knew exactly whether a particular test would come back always negative, always positive, or either way. But all 27 tests were systematically administered to all patients regardless of relevance, which wasted the patients' time, inflicted unnecessary suffering, and wasted hospital resources. The analysis revealed that, by going down the decision tree, some diseases could be diagnosed with only 4 tests, while the average for all diseases was 8. The generation and use of such decision trees, however, requires special software implementing an information-theory-based technique called "induction."

The idea is to replace the above figure of merit with a new one defined as:

$$\text{Figure of merit (step)} = \frac{\text{Information gain}}{\text{Time required}}$$

or

$$\text{Figure of merit (step)} = \frac{\text{Information gain}}{\$\ \text{spent}}$$

Testing multiple units at once

Sometimes it is possible to increase the productivity of test operations by testing multiple units at once. One such method was discovered in World War II in testing soldiers' blood for syphilis. It consisted in mixing the blood samples from 10 patients and testing the mixture. If the result is negative, then none of the patients is infected. If any of them is infected, then the result is positive, and all 10 are recalled for individual tests.

If the disease affects 5% of the population, then a mixed sample of 10 is clean 60% of the time. The other 40% undergo 1 test as part of the mix, plus one individually. To test 10 patients, we therefore need 1 test 60% of the time and 11 tests 40% of the time, for an average of 5.4, which is 46% less than it takes if we directly test them individually.

In cosmetics manufacturing, US Food and Drug Administration (FDA) rules mandate biological testing on liquid ingredients after 3 months in storage. The same logic can be used there as for blood tests. Without loss of accuracy, we can reduce the number of tests by mixing samples from several batches and even several ingredients.

This approach has also been applied to testing simple electronic components. The units to be tested are plugged into a series of outlets that pool their output signals. If any of the units malfunctions, all are retested individually. The down side of this approach for this kind of testing inside an assembly line is that it is a batch process and that the amount of operator intervention required varies depending on whether individual retesting is needed.

Overall, there are two major limitations to this approach:

Its value greatly decreases as the proportion of defective units in the population increases. If it is 20% instead of 5%, then a group of 10 will pass only $0.810 = 11\%$ of the time, and the average number of tests for a group of 10 will be $1 \times 11\% + 11 \times 89\% = 9.9$, which is essentially the same as if we had tested all group members individually, but with lower administrative efforts.

3 The batching and variability of the operator workload is a major drawback for operations taking place on an assembly line, which limits the approach to offline testing, as in the above example of cosmetics ingredients.

17.2.5. **Quality control labs**

In the process industries, it is common to organize testing in separate labs that are under the responsibility of the Quality department. They are common in food and beverage, pet food, paint, chemicals, pharmaceutical companies, among others. For example, after a mixing process, a sample is taken from the batch and analyzed. The batch is on hold during testing. If the sample is accepted, the batch is released for the next process. If the sample is rejected, it must be reworked or scrapped. Samples are carefully labeled and stored for years to enable tracking batches in case of quality errors, or to comply with external mandates.



Figure 17.6 A quality control laboratory testing pet food

In pharmaceuticals it is not uncommon that the quality control lab is located far away from the factory, even on another continent. This is because of the need for particular expertise or equipment and the

need to comply with external mandates – such as, for example, those from the FDA for all drugs sold in the US. The quality control lab procedures are subject to many items of the long list of regulations listed by the FDA. Samples are shipped to the lab and tested for acceptance.[12]

> **Stop and Think 17.3**
> Why is quality control organized differently in process industries like food and pharmaceutical compared to discrete manufacturing like cars and phones?

17.3. Data-driven quality improvement

This section introduces Statistical Process Control, Design of Experiments, six sigma, and the application of modern data science to this field.

17.3.1. Statistical Process Control

When manufacturing professionals talk about Statistical Process Control (SPC) they usually mean *control charts* and similar techniques dating back from the 1920s to World War II. It is a sophisticated tool based on the statistical state of the art 100 years ago, cleverly working around the limitations of that era's information technology.

There are entire books just about this topic. The most recent types, the cumulative sum control chart and the X-mR chart, are from the 1940s. They were the best you could do with the ideas of that time, paper, pencils, slide rules, and books of tables. These charts show the evolution of *one* quality characteristic measured after the operation is complete and issue alarms when it shows evidence of a change with an assignable cause. There was, at the time these charts were developed, no other way to capture and analyze the vital signs of the process itself during the operation.

Figure 17.7 shows an example of the $\overline{X}$-R chart that Shewhart designed to monitor the accuracy and precision of a process, in terms of a variable measured on workpieces post-process:

- The $\overline{X}$ **chart** is a plot of a sequence of group averages, with the group index on the x-axis.
- The **R chart** is, likewise, the sequence of ranges of values in the groups.

The Upper Control Limit (UCL) and Lower Control Limit (LCL) bound the fluctuations – "*common-cause variation*" in the SPC literature – around the overall mean caused by the process operating under expected conditions. As long as the process stays within the upper and lower limits then the process is "in control." Values outside control limits on either charts are anomalies with *special causes* that need to be investigated. Also trend patterns that systematically shift the mean should be checked as it may lead the process to get out of control in the future.

12 Friedli, T. & Bastoen, B. (2006). *Operational excellence in the pharmaceutical industry*. ECV, Editio Cantor Verlag.

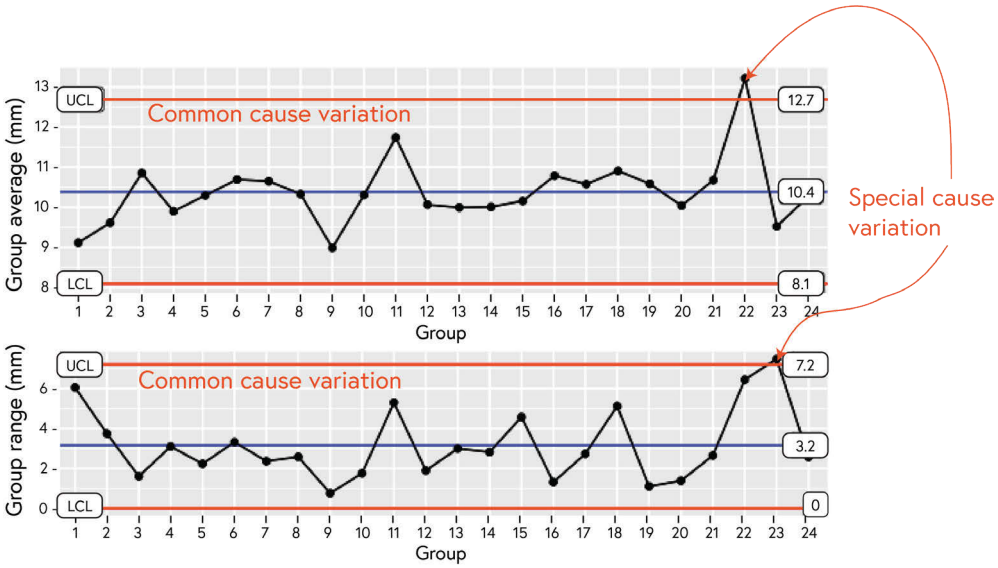


Figure 17.7 Vintage $\bar{X}$-R 1924 control chart

The x-axis, on a control chart, does not directly represent time but the sequence of parts that went through an operation. Two consecutive points represent two lots, or workpieces, processed consecutively, regardless of whether the time between them was 1 minute or 1 day, which may make a difference if the operation uses materials with expiration times, like paints or glues.

In the semiconductor industry, SPC has never been a major tool. The key parameters of circuits cannot be measured until testing at the end of a process that takes weeks and hundreds of operations, and the root causes of problems are often more complex interactions between features built at multiple operations than can be understood with the tools of SPC. In addition, the quantity of data generated is much larger than anything the SPC techniques were meant to handle. If you capture 140 parameters per chip, on 400 chips/wafer and 500 wafers/day, that is 28,000,000 measurements per day.

In modern manufacturing, SPC is obsolete for most processes. The SPC literature recommends measurements over go/no-go checking, because measurements provide richer information, but the quality problems in machining or fabrication today are often generated by discrete events like tool breakage or human error, including picking wrong parts, mistyping machine settings or selecting the wrong process program. The challenge is to detect these incidents and react promptly, and, for this purpose, go/no-go checking with special-purpose gauges on all parts is faster and more effective than taking measurements on samples.

In a nutshell, SPC is yesterday's statistical technology to solve the problems of yesterday's manufacturing. It doesn't have the power to address the problems of today's high technology, and it is unnecessary in mature industries. The reason it is not completely dead is that it has found its way into standards and training programs. This is why you still see control charts posted on hallway walls in many plants. Computers handle much more data and more complex relationships among them than what can be shown in a control chart but didn't exist in the 1920s. SPC is both obsolete and difficult to master.

17.3.2. Statistical Design of Experiments

Next to SPC, Design of Experiments (DOE) is the most common topic in discussions of statistical quality. In manufacturing, DOE is intended to help engineers design products and improve processes. Mastering DOE is a multi-year commitment that few manufacturing professionals have been willing to make. Furthermore, its effective use requires DOE know-how to be combined with domain knowledge. Outside of niches like semiconductors or pharmaceuticals, however, there is little evidence of its use in manufacturing.

What is DOE?

In DOE you choose the observed variables and collect data specifically for the purpose of solving a problem. You plan every step from data collection to the presentation of results. You decide which variables to collect, by what means, in what quantities, how you will analyze the results, what thresholds will make you conclude one way or the other, and how you will justify your conclusions to stakeholders.

Designed experiments yield high-quality data in quantities ranging from a few tens in small engineering problems to tens of thousands in the later phases of clinical trials for new drugs or A/B testing of webpage designs. In manufacturing, data are usually "small." Mathematician Ronald Fisher[13] was explicit in his focus on small data when he wrote: "Only by tackling small sample problems on their merits […] did it seem possible to apply accurate tests to practical data." To draw inferences from small samples, you need thresholds to mark differences as "significant." You test the observations against a null hypothesis, according to which there is a low probability that they should fall out of a given interval. You set this probability, or level of significance, arbitrarily set at 5%, 1%, 0.3%, or 3.4 ppm. "Significant" means nothing unless you specify the level.

DOE comes into play when results are not obvious but answers are needed. It originated in agricultural research where a fertilizer may enhance crop yield by 5% more than another but where such differences have tangible economic stakes. When you treat the same crop with different fertilizers in different plots, you observe different yields. Figure 17.8 shows an arrangement of plots used at Rothamsted, UK, for wheat growing experiments, ongoing since 1843. Any result doesn't, however, *prove* that one fertilizer works better than another. You work with small datasets because collecting data takes years. The differences you observe may be fluctuations or may be due to other factors that agricultural science cannot explain. Fisher, who was not an agronomist but a mathematician, developed methods for assessing the reality of these differences.

The lesson from Fisher's book is not that there is a one-size-fits-all methodology for statistical DOE. Instead, you tailor it to the needs of each domain. We still use many of the tools he developed, particularly *analysis of variance* (ANOVA). Today, Google uses the simplest of Fisher's techniques when they do *A/B testing* on webpage details that may be as small as the color of a confirmation button. You choose users to receive one version of the page or the other as if they were different fertilizers. Then you harvest the resulting crops of clicks. You don't rely on any science of color preferences. You assume that the color of the button makes no difference, and analyze the data for evidence of the contrary.

13 Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd.



Figure 17.8 The Rothamsted Broadbalk field for crop growing experiments in the UK (Credit: Rothamsted Archives)

These kinds of *one-factor-at-a-time* (OFAT) experiments are simple and easy to understand but often insufficient. You can use fertilizers in different doses, combine different ones, and irrigation patterns also play a role. In addition, crop size may not be the only output of interest. Quality matters too. And when there are interactions between factors, you cannot reach the best combination of levels by OFAT experiments. As data collection is time-consuming and expensive, you end up with small datasets for each combination of levels.

In his summary of DOE for manufacturing, Viktor Astakhov[14] uses Figure 17.9 to explain the general context:

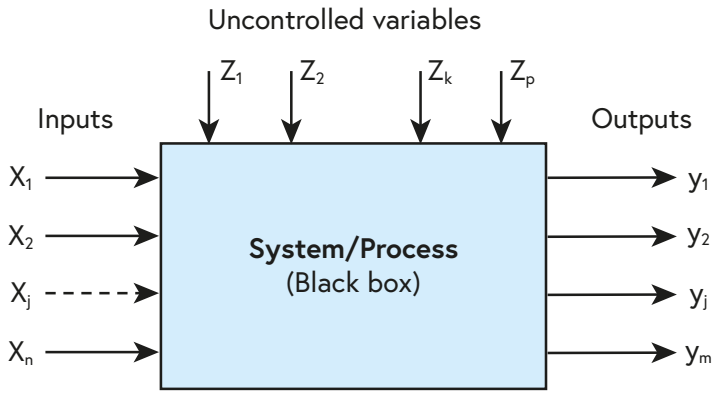


Figure 17.9 The object of DOE in manufacturing (Credit: Viktor Astakhov)

DOE aims to correlate outputs $(y_1,...y_m)$ with inputs $(x_1, ..., x_n)$ in the presence of noise $(z_1,..., z_p)$ and in the absence of a mathematical model of the system or process. The experiment, or the sequence of experiments, lets you build an empirical model. It relates outputs to inputs without opening the black box. We still don't know what happens inside but we have a model of its consequences.

14 Astakhov, V. P. (2012). Design of experiment methods in manufacturing: Basics and practical applications. In *Statistical and computational techniques in manufacturing* (pp. 1–54). Springer.

While the DOE model looks general, it doesn't fit all the experiments scientists and engineers do. They often don't start with a black box but with a generic model of the system and they experiment for the purpose of assigning values to model coefficients.

When Fisher was pursuing the highest crop yields, he only cared about aggregate quantities for plots. Variations between individual plants within plots were not a concern. It's different in manufacturing, where the goal is to produce identical units of the same item. In manufacturing, the challenge is to first reduce outcome *variation*, and then to improve outcome precision.

Factorial experiments and Taguchi's robust design

Astakhov illustrates DOE with a cake tasting experiment having four input factors:

- Oven temperature
- Sugar quantity
- Flour quantity
- Number of eggs

With only two settings for each, that's $2^4 = 16$ combinations of settings to bake cakes with, randomizing the sequence as needed to filter out the influence of factors that are not part of the experiment, like the chef or the carton the eggs come from. Then you use the data to assess the influence of each factor and of interactions between factors. You can group the data by any combination of factors, compute Fisher's ratio of the variance of the target output between and within groups and, if it is deemed significant, use the group means to estimate the output produced by this combination of factors.

In a full factorial experiment, you generate data for all 16 combinations of settings and are able to analyze all the factors individually, in pairs, in triplets, and all four together. That's all the non-empty subsets of the set of factors, of which there are $2^4 - 1 = 15$. While this is work, it is considerably less than running separate experiments for each factor individually and for each pair, triplet, etc.

The work of doing full factorial experiments rises quickly with the number of factors and the number of levels considered for each factor. If you increase to 5 factors with 3 settings for each, the number of combinations for which you need to generate data rises to $3^5 = 243$, to evaluate the effect of $2^5 - 1 = 31$ combinations of factors. This is why you screen to limit the number of factors to consider. You also need to keep the number of levels or settings for each factor low. Often, you end up with partial factorial designs that include only some of the possible combinations of settings.

The idea is to take advantage of the usually decreasing effect of combinations of factors when you include more. Individual factors contribute more than pairs, which contribute more than triplets, etc. The two oven temperature settings alone will account for more of the variance in the result than combinations of settings for oven temperature, sugar quantity, flour quantity, and number of eggs. You take advantage of this to generate partial factorial designs when adding new factors.

Genichi Taguchi popularized his concepts in the US in the 1980s and they are still referenced as the state of the art for DOE applied to manufacturing quality. He focused on the elimination of *variability*. Some of Taguchi's concepts are straightforward, like the quadratic loss function but others are dauntingly complex, which may explain why they are not more widely used. An alternative and more

descriptive name is "robust design." The key idea is to search for the combination of controls that is least sensitive to noise.

Taguchi's quality loss function

The idea behind tolerances is that, if the critical characteristics of a product unit all fall anywhere within an interval, the unit works. If there are 45 critical characteristics, their tolerance intervals form a 45-dimension hypercube. The actual space of critical characteristics for which the unit works can be of any shape and the target is to make the tolerance hypercube small enough to fit within that shape, yet not so small that it is beyond the capability of the process.

If you apply this logic rigorously, your loss is 0 if the unit is anywhere inside the cube and constant outside of it, because you reject the unit anyway. It doesn't matter whether you missed the interval by an inch or a mile. With his quadratic loss function around a target value, Taguchi's loss function shifted emphasis from aiming for characteristics within a tolerance interval to aiming for a target value.

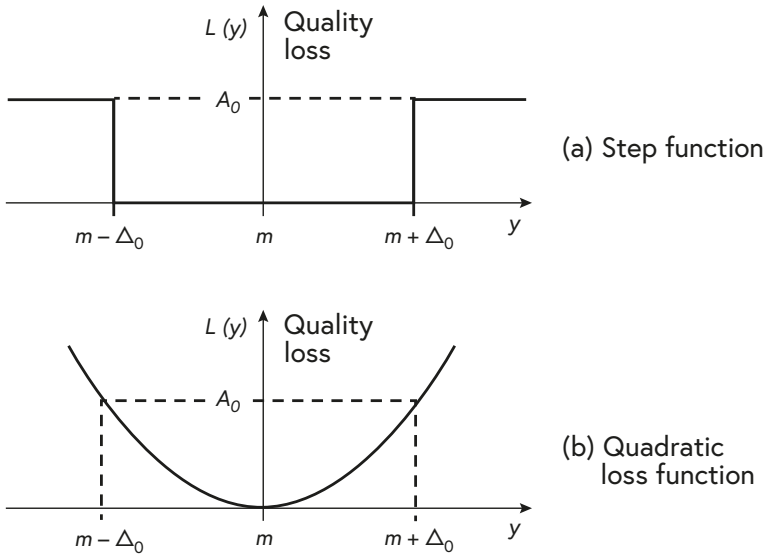


Figure 17.10 No loss within tolerance interval versus Taguchi's quadratic losses

Taguchi splits the inputs into *signal factors* and *control factors*. The signal factors are the user settings on a product or a machine when running it. The control factors are parameters set at design time. They are the characteristics of the hardware and software it is built from. The *noise factors*, as before, are outside the control of the user or the designer.

The objective of the Taguchi experiment is to identify the control factors so as to maximize a "signal-to-noise ratio" defined as the square of the coefficient of variation of the output, expressed in decibels. This idea has not been universally accepted by statisticians.

17.3.3. **Six sigma**

Motorola launched six sigma in 1987 with the objective of eliminating defects and process variation. The goal was to achieve "six sigma performance," where sigma refers to the standard deviation is a normal distribution. A *six sigma performance* means that at most 3.4 defects per million opportunities (dpmo). What it translates to in terms of defective parts per million (dppm) depends on the number of defect opportunities of the parts. An automatic transmission for cars has ~2,200 critical dimensions, each of which is a defect opportunity. In this case, 3.4 dpmo translates to 7452 dppm, which is not a particularly high level of quality.

The 3.4 dpmo threshold does not actually correspond to fluctuations around a constant mean with a standard deviation σ with limits at $\pm 6\sigma$. Instead, the process mean is assumed to fluctuate within $\pm 1.5\sigma$ and the limits are set 4.5σ beyond this interval on both sides. This model has been largely dismissed by statisticians as arbitrary.

In 1992, six sigma programs was extended with an organizational structure, including improvement projects facilitated by internal consultants called "Black Belts" and many organizations started offering training for certifications as yellow, green, and black belt, with a more senior level called Master Black Belt (MBB). ISO published its first certification for six sigma in 2011. It represented a break with TQM, because it wasn't about participation by everyone.

However, the *belt system* has been particularly successful as a certification scheme for skills and, therefore, flourished. Having a colored six sigma belt can advance careers and are sometimes requirements listed in job descriptions in manufacturing, independently from whether or not it results in real improvements.

The names most commonly cited as creators of six sigma are Bill Smith and Mikel Harry.[15] Initially, at the insistence of Allied Signal's Steve Zinkgraf and Gary Cone,[16] the technical content of six sigma included statistical Design of Experiments (DOE). Over the years, however, the technical content of six sigma withered back to SPC. Nevertheless, six sigma was widely adopted.

At General Electric (GE), CEO Jack Welch imposed it during his tenure, going as far as requiring Black Belts for promotion to management, which led GE employees to package all their activities as "six sigma projects." Starting in 2001, six sigma at GE was gradually discontinued by Welch's successor, Jeffrey Immelt, and not revived by Immelt's successor, Larry Culp, who took over in 2018.

In 2003, the six sigma community sought to strengthen the brand's appeal by coining "lean six sigma" as a broader term, incorporating content from the lean literature. The main legacy of six sigma is the belt system and the generic problem-solving framework *DMAIC*, which stands for define-measure-analyze-implement-control. While six sigma has withered, it is still taught by many consultants and in some academic programs. It remains popular with many engineers, particularly those who earned a belt in it.

15 Harry, M. & Lawson, R. (1990). *Six sigma producibility analysis and process characterization* (Publication Nr. 60-3-03/88). Motorola University Press.

16 Zinkgraf, S. A. (2006). *Six sigma - The first 90 days*. Pearson Education.

17.3.4. Data science for quality improvement

In manufacturing today, *data science* is about finding information buried in a tangle of inconsistent, inaccurate, and archaic legacy systems combined with data collected from machines and retrofitted IoT applications. It is not a substitute for going to the shop floor to directly observe operations and listen to what the people have to say but it is a complement to it. It tells you the backstory to what you can see, and it validates what people tell you. Most powerfully, it can even help you see things you would not be able to discover yourself. When working to apply data science to improve manufacturing operations, it is essential – and easy to forget – that the starting point is not technology but the people who record, store, retrieve, crunch, interpret, and use the data.

High-technology industries like semiconductors exist in a world where products are obsolete when processes are mature. As a consequence, they are still pushing the state of the art in using data to develop and sustain their process capabilities, with methods that are not acknowledged or taught by the quality profession. The key word in these industries is no longer statistics but *data science*.

What is data science?

The discipline of *statistics* has been subsumed under *data science*. Statistics is about mathematical methods to draw conclusions from data. Data science has a broader scope: It starts with data acquisition and ends with finished data products that can be reports, dashboards, or online visualizations and query systems. Since analyzing complex relationships in data requires computers, data science is closely linked with computer science.

When hearing "data science," many think of artificial intelligence (AI). AI as an analytical tool is an important part of data science but there is more to it. Machine learning (ML) is a subdiscipline of AI, and is representing algorithms that become better at a task with repetition. For example, a neural network may be designed to recognize a handwritten "8" and to improve its performance with experience.

Data science needs data

Data about operations can be split in terms of subject and relationship to time. Within a manufacturing company, for example, the *subjects* include production, engineering maintenance, production control, logistics, supply chain management, warehouse management, quality management, accounting, human resources, etc. With respect to *time*, the data breaks down as follows:

- **Timeless.** Master data that changes when management decides to change it, not as a result of the passage of time or events.
- **Future.** Plans, schedules, forecasts, or bookings are data about the future that reflect the organization's expectations and commitments. This data changes as the future becomes the present.
- **Present.** This is data about current status, like what the organization is currently working on, the resources available, and its backlog. It changes as a result of events like job completions or machine failures and when materials reach their expiration date or a maintenance action comes due.
- **Past.** Historical records are kept for traceability and to support analytics. Once stored, they don't change, and new ones are appended constantly.

You have to start from the data you have and it is usually not pretty. You have to check its quality, in all its dimensions, correct errors and omissions where possible, and flag the records where it is not. You extract the information encoded in names and decrypt it. You review missing values to assess whether they are for a parameter that is not applicable to particular records, are deliberate omissions, or are meaningful. Then you have to format the results as valid input to analysis. Data scientists spend more of their time preparing data than analyzing it but get little credit for it from statisticians, who are trained to assume a clean dataset as a starting point.

The case for explainable AI

Explainable AI (XAI) is a set of methods that support humans in understanding how AI models map certain inputs x – for example, machine measurements, time stamps, product images – to certain outputs y – for example, probability estimates of yield, safety performance, schedule accuracy. Data scientists such as Cyntia Rudin separate between interpretability and explainability:[17]

- **Interpretability** is the ability of models to be analyzed to understand their behavior. Decision trees, for example, are interpretable.
- **Explainability** is the predictability of the output of an AI black box by simpler AI algorithms that "explains" the behavior of the black box.

Hence, explainable AI refers to a set of methods that reverse engineer the prediction of a blackbox AI model by, for example, using game theory. This is useful because the most advanced AI models – "blackbox" deep neural network models or gradient boosting machines – are not *interpretable* but *explainable*.[18] We prefer an interpretable model over an explainable one as long as its accuracy is similar or better but this is not always the case. In manufacturing, we would like to have interpretability or explainability for several reasons:

- **Trust.** If engineers, operators, and managers are going to trust the output of AI, they must be confident that the AI model is accurate.
- **Follow-up action.** If users are to act on the output, they need to know what to do. Explainable AI can point to a process or measurement that engineers can investigate further.
- **Decision power.** AI decision-support. Only for low-risk processes, where AI has been shown to have high performance, should the processes be fully automated. An example is visual inspection of work-in-process. In all other cases, domain-experts should be able to overrule the AI output if they know it is wrong, for example because they possess important information that is not in the AI model.

Box 17.3 shows an example of explainable AI used for root cause analysis in semiconductors. The example is from Hitachi ABB, today Hitachi Energy,[19] which uses data science to learn the underlying relations among the many variables affecting semiconductor yield. This way, they can pick up factors that are neither visible to the human eye nor detectable with SPC or DOE. In addition, once appropriate data are available and formatted, the analysis is fast and easy to run.

17 Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*, 206–215.

18 Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR), 51*(5), 1–42.

19 Senoner, J., Netland, T., & Feuerriegel, S. (2022). Using explainable artificial intelligence to improve process quality: Evidence from semiconductor manufacturing. *Management Science. 68*(8), 5704–5723.

Box 17.3 **AI for root cause analysis**

AI can detect hidden relationships in manufacturing settings with thousands of features. At Hitachi Energy, chips are fabricated on silicon wafers through 200 operations carried out in a clean room in batches of 25 wafers. After a batch is completed, each wafer is cut into 62 chips. The chips are then tested and yield is the ratio of passing versus failing chips (Figure 17.11). Some chips fail. Why?

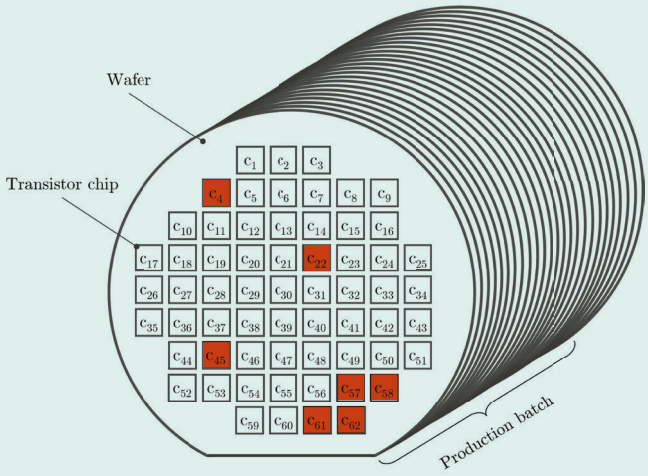


Figure 17.11 What causes yield loss? This wafer has 7 defective chips out of 62, or a yield of 88.7%

It is not easy to find the causes of yield losses in semiconductor fabrication. Each operation has many parameters and settings interacting with each other and with those of other operations in non-linear ways. Semiconductor manufacturers collect several thousand measurements for each chip. SPC falls short. The common practice is a combination of linear regression and DOE, where domain experts test hypotheses of possible root causes for yield loss. But DOE is expensive, time-consuming, and also error-prone in environments with complex interactions.

Hitachi Energy combines deep learning with the *SHAP value method* to infer the prediction rules. SHAP decomposes a prediction into contributions from each predictor. The SHAP value of each feature is plotted against its strength and contribution to the overall variation in yield. A SHAP summary plot gives decision support to yield and process engineers who look closer to eliminate the root causes of yield loss.

Figure 17.12 shows a worker using a visual inspection system with explainable AI from EthonAI. The photo on the touchscreen is overlaid with a heat map that "explains" to the operator if and where the AI system flagged a defect. This explanation helps the operator trust the AI system, since he or she can immediately verify the suggestion by inspecting the real part. It also reduces the mental load and time the operator needs for finding defects during the inspection.

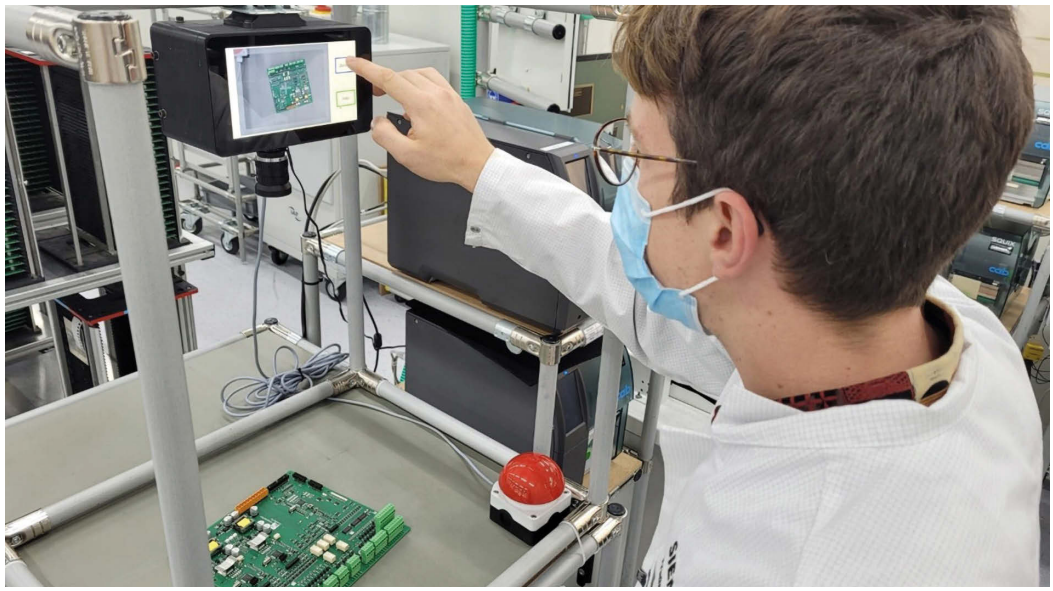


Figure 17.12 Explainable AI used in a quality control task with visual inspection (Credit: EthonAI)

17.4. Built-in quality control

Radically different from trying to identify and fix defects that have happened, the alternative strategy is to prevent defects happening in the first place. If your process produces 30% of defectives, this is out of reach and you need to focus on process capability. Once the process is *capable*, there are ways to engineer the products and processes for defect prevention and containment. These methods directly work on shop floor operations, affect all parts, and do not involve data science. In spite of their effectiveness, they have yet to be acknowledged by the quality profession.

17.4.1. Quality implications of one-piece flow

The one-piece flow pattern is shown in Figure 17.13 using a cell as an example. From the time it leaves the input buffer to the time it reaches the output buffer, each part moves individually from step to step *without any accumulation of WIP* between machines or work stations.

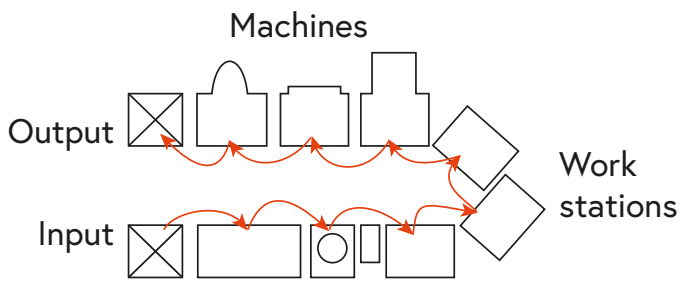


Figure 17.13 One-piece flow through a cell

This has several consequences for quality:

- The lead time through this sequence of operations is an order of magnitude shorter than with batch and queue between every two operations. If parts are tested after coming out, negative test results can be fed back to the line before it has a chance to damage more than a handful of parts.
- Because parts move first-in-first-out (FIFO) at every step, the process sequence is preserved, and the parts affected by any defect detected at testing can be identified, as shown in Figure 17.14.
- Within the process, any defect introduced on one part at a step that prevents successful execution of the next step is detected immediately and production of further defectives prevented.

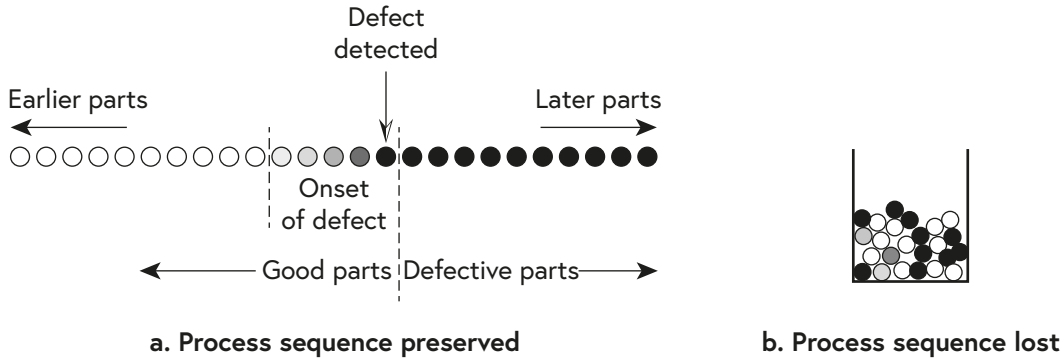


Figure 17.14 Quality impact of preserving process sequence

The effects of one-piece flow and FIFO are backed up by in-process quality checks, which take the form of go/no-go gauge checks or of what Shigeo Shingo called "successive inspection" in assembly, where each assembler touches the parts installed by the previous one. Figure 17.15 shows examples of go/no-go gauges.

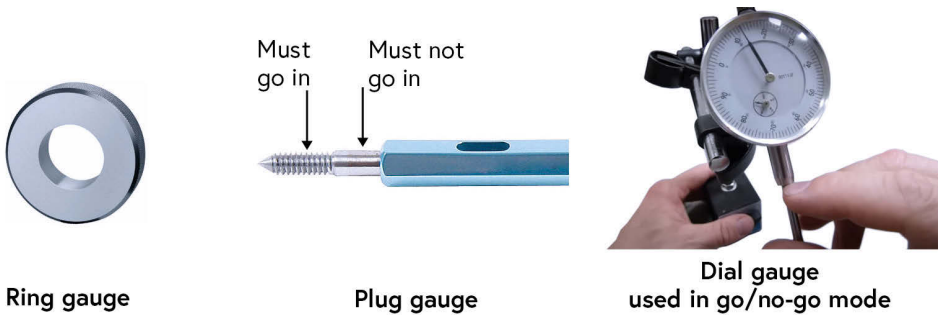


Figure 17.15 Examples of go/no-go gauges

For any characteristic, the options are:

1. **No checking.** If a particular dimension has never been found out of spec for 10 years, it probably doesn't need to be checked.
2. **First and last piece checking.** It is applicable to production runs where the characteristics for parts in-between can be interpolated. The checking operations then become part of the setups rather than of production tasks.
3. **100% checking.** As part of their routines, operators check every part, and place defectives in a red box where it is collected by the Quality department.

These concepts are difficult for quality professionals trained on SPC to accept because they exclude both measurements and sampling. They have been trained to think that you should never use a go/no-go gauge where you can measure, because a measurement is richer information. The measurement tells you more than whether a dimension falls between goal posts; it tells you where. They have also been told that measuring 100% of the parts is unaffordable, and that well-designed sampling plans can give you the same information at a lower cost.

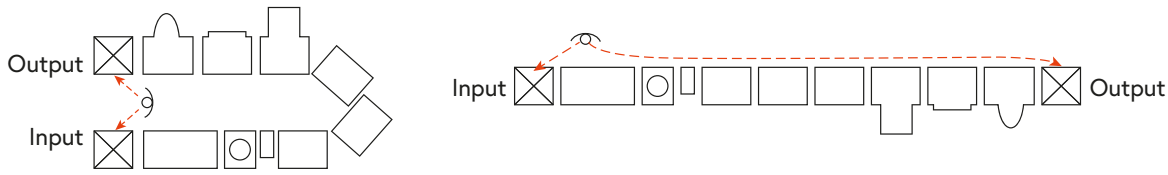
Sampling is not allowed inside cells, for several reasons:

- **It allows defective parts to wait and accumulate between measurements**, which does not serve the objective of detecting and responding to problems quickly.
- **It disrupts operations.** Operator jobs are choreographed into a sequence of motions that is repeated with every part. Add measurements to be taken on every fifth part at one operation and every seventh part at another, and you throw off the operators' routine. On the other hand, a check on *every* part can be made a part of this routine, and go/no-go gauges can be engineered to make these checks fast enough not to slow down production.

Also, given that the problems we are trying to detect are discrete state changes rather than drifts or fluctuations, the information loss due to using go/no-go checks instead of measurements is not as great as it would have been with process capability issues, and is more than made up by application to every part rather than to a sample, and by faster problem detection.

As noted earlier, the compilation of measurements and attributes into control charts is done to satisfy external mandates. This can still be done without disrupting the flow of production, for example, when parts are finished or while they are between segments of the process that have organized into cells. Measurements, of course, also have a place in the solution of problems once they have been detected.

One-piece flow and FIFO do not imply and are not always implemented in the form of a U-shaped cell. The U-shape, however, also has an impact on quality, simply because seeing incoming and outgoing parts side-by-side makes operators more conscious of quality issues. This is best understood by considering the opposite, as shown in Figure 17.16.



a. Operator sees the result of the team's work **b. Operator cannot see the result of the team's work**

Figure 17.16 Quality impact of contiguous input and output buffers

17.4.2. Self-inspection versus inspection

The TPS approach, as described by Shigeo Shingo,[20] is to avoid sampling as a disruption of production routines, and focus instead on checking 100% of the units but doing it in a way that does not slow down production. This means first using go/no-go gauges and moving on to mistake-proofing to eliminate that step. Also other books on TPS, like Monden's,[21] give the subject short shrift by seeing inspections as waste in assembly work. TPS-proponents claim that quality must be *built into the process*, and each assembler is responsible for the quality of his or her work. In this view, inspections are sin and must be done away with. Returning to the shop floor, you are confronted with the fact that these activities cannot be done away with in the short term, and possibly even in the long term.

Specialized inspectors are not only additional labor, but an additional labor category, whose work needs to be coordinated with production to ensure that an inspector is available when the output of a production operation is ready. In daily operations, it is simpler if assemblers can inspect their own work.

Quality professionals, however, object to "self-inspection" on the grounds that workers do not have the objectivity required to inspect their own work, and this is a valid concern. In Juran's handbook, Joseph Zeccardi describes the purpose of inspections as "determining whether product conform to specification."[22] While this translates into "making sure the product is right," this isn't the mindset that an inspector should have. Instead, an inspector should always assume there is something wrong and take on the mission of finding out what it is. Inspection or test productivity is measured in terms of defects found per unit time or per quantity produced, and a unit that passes is an inspection or test "failure."

The need for this negative attitudes towards the product causes an organization dilemma:

- A worker who has just built a unit cannot turn abruptly into a ruthless investigator of its defects.
- A separate corps of inspectors develops an adversarial relationship with production, in which inspectors assume that production is deliberately trying to get bad units past them, and in which they end up celebrating "big busts" when they uncover a problem.

There are several solutions:

1. For manual assembly, successive inspection is a viable alternative to self-inspection. Instead of trying to find defects in his or her own work, each assembler does it on the work done at the previous station. This usually provides enough psychological distance for it to be possible.
2. For mechanized or automated assembly, the work of putting components together is not done by the assembler but by a machine, and self-inspection is possible, because it is the work of the machine that the assembler checks.
3. The clearer the inspection criteria, the easier self-inspection is. The questions of whether all screws are in place, a plug gauge goes into a hole, or a green light on an instrument comes on, leave no room for interpretation or subjectivity. The same is not true, for example, of inspecting a painted surface for scratches or runs.
4. What is most difficult for an assembler to do is design a self-inspection procedure. On the other hand, it is not nearly as challenging to go through a checklist designed by someone else.

---

20 Shingo, S. & Dillon, A. P. (1989). *A study of the Toyota Production System: From an Industrial Engineering viewpoint*. Productivity Press.

21 Monden, Y. (2010). *Toyota Production System: An integrated approach to just-in-time*. Productivity Press.

22 Zeccardi, J. (1988) Inspection and test. In: Juran, J. M. & Gryna, F. M. (Eds.) (1988). *Juran's quality handbook: The complete guide to performance excellence* (4th ed.). McGraw-Hill Education. Chapter 18.

Companies often neglect to provide clear inspection criteria. They may produce "quality plans" that fill binders while failing to let inspectors on the floor know exactly what they should be looking for and leaving acceptance of parts to their judgement. Nothing is more likely to stir up conflict between an inspector and managers than a vague spec.

Other issues with self or successive inspection include the additional training required and the production lost while operators inspect instead of making. Both of them are addressed in the design of the inspection. Easy-to-apply go/no-go criteria both simplify training and reduce the time required. The productivity issue must also be considered in light of the alternative or either taking the product off-line to an inspection station or waiting for an inspector to come to the assembly station, either of which will slow down the flow of the product more than self or successive inspection.

17.4.3. **Mistake proofing**

Shigeo Shingo[23] explained how he coined the term "poka-yoke" in 1963 after an operator was offended by the word "baka-yoke," a common everyday word for "foolproofing." The operators are no fools but everybody makes mistakes. Poka-yoke translates to "mistake-proofing," which has been adopted into quality standards documents like QS-9000 and in the lean literature.

Besides Shingo's book, the main source of information on mistake-proofing in the US is the big red book,[24] a compilation of 240 examples from 1980s Japanese industry, presented in the form of before-and-after sketches with comments. Poka-yoke is best learned by example. The "big red book" contains little theory. In the 2020s, however, many of these examples are dated. C. Martin Hinckley's *Make No Mistake!*[25] is a more recent collection of examples presented in the same format. In Japanese, several authors have regularly updated the field of poka-yoke, particularly Hiroyuki Kadoma.[26]

Hinckley's poka-yoke examples include visibility methods that make mistakes *unlikely*, which is valuable but falls short of making processes mistake-*proof*. The discipline of designing products to make user errors unlikely is "*usability engineering*." Don Norman[27] explains how to design objects like door handles, light switches or faucets in a way that leads people to use them as intended with little or no training. While short of mistake-*proofing*, these concepts are applicable to the design of operator-machine interfaces and sometimes sufficient. If the visual methods are photos and drawings, we call them visual instructions. We reserve the term mistake-proofing, or poka-yoke, for devices that either make mistakes impossible to commit or detect them as soon as committed.

> **Stop and Think 17.4**
> Can you think of any poka-yoke in your daily life?

23 Shingo, S. (1986). *Zero quality control: Source inspection and the poka-yoke system*. Productivity Press.

24 Nikkau Kogyo Shimbun (Ed.). (1988). *Poka-yoke: Improving product quality by preventing defects*. Taylor & Francis.

25 Hinckley, C. M. (2001). *Make no mistake! An outcome-based approach to mistake-proofing*. Taylor & Francis.

26 Kadoma, H. (2019). *Poka-yoke countermeasures: Preventing human error! Human error countermeasures*. Monozukuri Library.

27 Norman, D. (2013). *The design of everyday things*. Revised and expanded edition. Basic Books.

Figure 17.17 shows two examples contrasting usability engineering and poka-yoke



Figure 17.17 Poka-yoke examples

To the left is a person filling gasoline on a car. The different colors of fuel pumps for gasoline and diesel help people pick the right one but don't ensure that they do, hence it is usability engineering. However, if the person has a gasoline car and selects the diesel pump, fueling will be prevented because the nozzle of the diesel pump is too big to fit into the petrol filling hole. Unfortunately, this design only works one way: It keeps the customer from filling a gasoline tank with diesel fuel but not the other way around. It lets you fill up a diesel tank with gasoline, which has worse effects on the engine than the opposite mistake. Hence, it is only a poka-yoke if you drive a gasoline car. The second example is a mechanical wrench with automatic torque stop and "click" feedback. It is a technical tool that prevents operators from overtightening screws and ensures that the specified torque value is applied.

The following common characteristics emerge from when you consider poka-yoke examples:

1. **Poka-yoke being devices are all about people.** A feature in an automatic machine that detects and responds to malfunctions, while useful, is not a poka-yoke. In a manual operation, the best poka-yoke makes a mistake physically impossible; in a machine operation, it prevents the operator loading the wrong parts or starting the wrong process. The second-best poka-yoke prevents passing on parts with defects introduced by the machine.
2. **A poka-yoke works 100% of the time.** It flows logically from what it is. A device that reduces the rate of human error by 95%, while very useful, is not a poka-yoke. A device that fails good parts, even rarely, is not a poka-yoke either.
3. **A poka-yoke does not add work for the operator.** This is essential, because devices that do add work are disconnected or bypassed under pressure.

Making poka-yoke solutions is a fun task, and something that should largely be left to the operators. Making the solutions themselves, operators are likely to be sure they work and will not find workarounds.

Many poka-yoke are simple, but not all. If implementing it requires modifying a fixture, improving a control program, or modifying the design of a product it is beyond the scope of what a shop floor team can do. It requires time, money, technical validation and management approval. The proponents of expensive solutions are usually asked to think harder, and sometimes with success. When, for example, two nearly identical bolts are frequently confused in assembly, rather than using a device to tell them apart, it can be easier to split the operation between two stations, each of which uses only one of the bolts. Box 17.4 shows two examples of poka-yoke for stud screws in an automotive factory.

Box 17.4 **Preventing installation errors for backward studs**

As shown in Figure 17.18, the asymmetrical stud, found in car engines, challenges assembly to prevent installing it backwards. The engineers' first solution was equipment with sensors, which was both expensive and a source of downtime. Plating dust from the studs contaminated the sensors and caused the flipping device to jam.

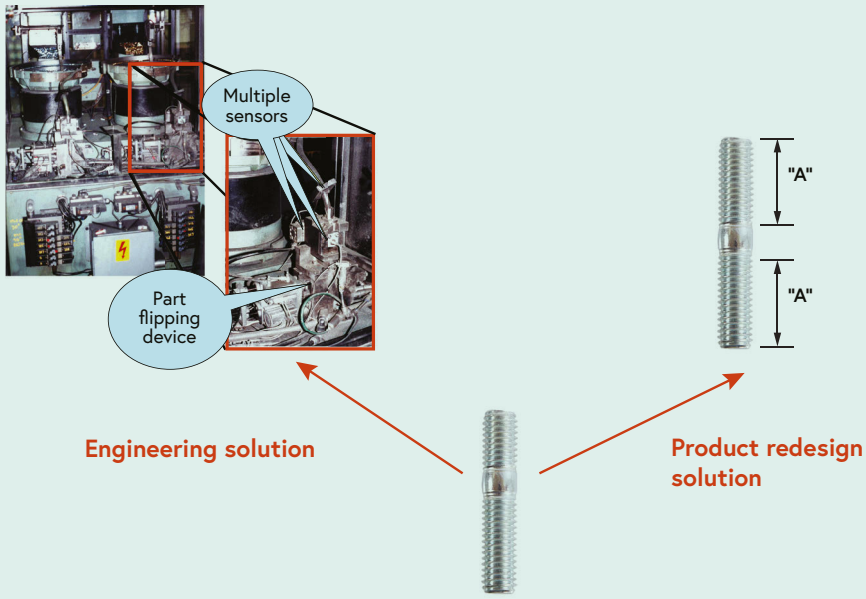


Figure 17.18 "Solutions" to the asymmetric stud problem

One day, a shipment arrived contaminated with parts intended for Toyota, that were symmetric and, as such, had no "backwards." While a more elegant solution, it was neither cheaper nor faster because it involved a change in the product design. They made a note of it for the design of the next product but couldn't apply it to the current one. In the meantime, they could use a concept like Kadoma's bolt length poka-yoke in Figure 17.19:

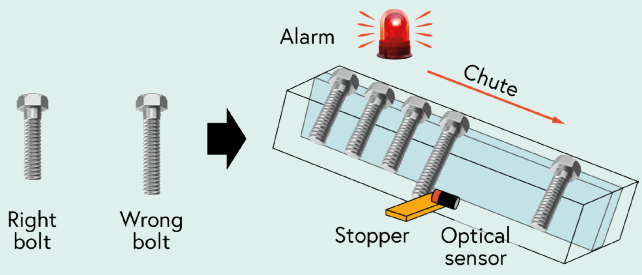


Figure 17.19 Kadoma's bolt length poka-yoke

17.5. Problem-solving frameworks and methods

When defects happen, we need frameworks to deal with problem-solving in a systematic way. This section introduces the PDSA or PDCA cycle, other problem-solving frameworks, quality control circles, Quick Response Quality Control (QRQC), and tools for root cause analysis.

17.5.1. Plan-do-study-act

The most consequential event in the quality field in this area may have been the lectures given by W. Edwards Deming in Japan in 1950, hosted by the Japanese Union of Scientists and Engineers (JUSE), known as Nikkagiren (日科技連) in Japan. According to Peter Kolesar,[28] the lectures were focused on control charts, and, at the end, acceptance sampling. As part of his introduction in the first lecture, he introduced the cycle shown in Figure 17.20, which is known as the plan-do-study-act (PDSA) method.

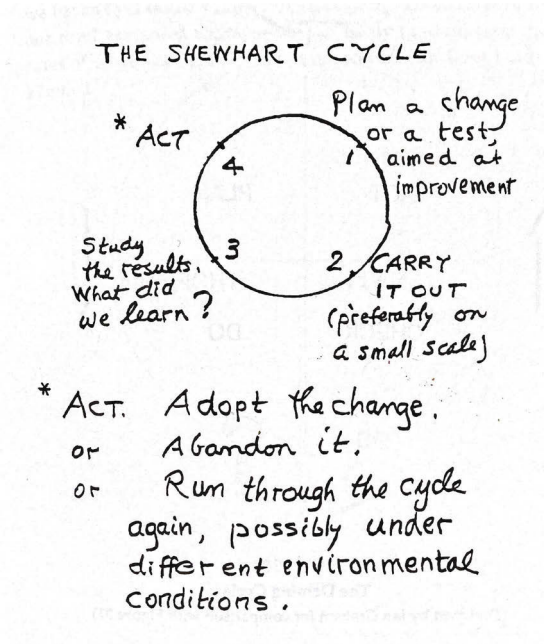


Figure 17.20 Deming's own PDSA drawing cited by Neave[29]

The idea caught on in Japan as PDCA instead, for plan-do-*check*-act Do and act may sound synonymous but, in this context, do means "execute the plan," and act means "correct or adjust as needed based on the results of checking." "Act" leads to the next iteration of "plan." PDCA is omnipresent in manufacturing today, but Deming himself devoted only two pages to PDCA in *The New Economics*[30] and made no mention of it in *Out of the Crisis*.[31]

28 Kolesar, P. (1994). What Deming told the Japanese in 1950. *Quality Management Journal, 2*(1).

29 Neave, H. R. (1990). *The Deming dimension*. SPC Press.

30 Deming, W. E. (2000). *The new economics: For industry, government, education*. MIT Press.

31 Deming, W. E. (1986). *Out of the crisis: Quality, productivity and competitive position*. Cambridge University Press.

Deming's Japanese students of 1950 applied his tools with great success in the following decades. Today, however, most companies have long stopped using them, except for PDCA, which has become pervasive, and not only in quality. Pascal Dennis,[32] for example, organized his account of Hoshin planning around PDCA. The Toyota way of working (Toyota shigoto no shikata, トヨタの仕事の仕方) is a refinement on PDCA taught to employees worldwide as "Toyota Business Practices" (TBP).

Box 17.5 **PDSA/PDCA in simulation games**

The LEGO Trucks game involves four 10-minutes rounds of assembling LEGO trucks, separated by 30 minutes for the teams to think through and implement improvements. Each run is a check on the last design. Within the 30 minutes between runs, each team reflects on what worked and what didn't, plans improvements to the assembly line, and implements these improvements. The whole game goes through three PDSA cycles.



Figure 17.21 A LEGO truck simulation session

The response of different players to this challenge vary. Some groups coalesce into teams and are comfortable with the disciplined approach; in others, forceful individuals attempt to push through their own solutions, with varying degrees of success; in others yet, individuals immediately make changes to the production line without consulting others.

PDSA is not as obvious as it sounds. The players come from all layers of the organization and, as a rule, play other roles in the game than in real life. Most of the players who struggle have been discouraged from doing anything at work other than follow orders. Others were managers and engineers steeped in a "ready-shoot-aim" culture that values nothing other than doing. They view planning as a waste of time and checking outcomes as a risk of exposing less than total success.

The teams that follow PDSA routinely outperform the others, and this is often a key takeaway for the players. While PDSA is trivial and obvious for some, it is worth emphasizing as a rational framework for improving processes.

32 Dennis, P. (2006). *Getting the right things done: A leader's guide to planning and execution*. Lean Enterprise Institute.

17.5.2. Imai's 7-step method, 8 disciplines, and DMAIC

Building on the same ideas as PDSA, authors and companies have suggested an array of more detailed structured problem-solving methods. The key point is to prevent humans under pressure from jumping to conclusions based on what Kaoru Ishikawa summarizes as "experience, intuition, and guts" (keiken-kan-dokyo, 経験. 感, 度強). They go by a solution that has worked before, they just *know* it applies to this new problem, and they must show the courage to act.

Applying problem-solving methods gives them a moment of pause and forces them to think through the issues before acting. What matters is not what you call the framework you use but how you solve problems. A structured framework helps not only solve problems more effectively and sustainably but also communicating about them with management, customers, and suppliers.

Masaaki Imai extends the PDCA model into a seven step problem-solving method:

1. Define the problem
2. Analyze the problem
3. Identify causes – for example by asking "Why?" five times or using fishbone diagrams
4. Plan countermeasures
5. Implement countermeasures
6. Confirm results
7. Standardize the new process

Another common framework is Ford Motor Company's 8 Disciplines (8D). Its objective is to "stop the bleeding" with immediate countermeasure, identify root causes using statistical methods, and eliminate recurring problems by implementing permanent corrective actions. The 8 disciplines are:

1. Build a team
2. Define the problem
3. Implement an immediate countermeasure
4. Determine the root cause
5. Verify the root cause
6. Implement a corrective action
7. Ensure prevention
8. Congratulate the team

Yet another framework is DMAIC from six sigma. DMAIC is a data-driven, customer-focused, structured problem-solving framework with an explicit focus on measurement. It stands for:

1. Define
2. Measure
3. Analyze
4. Implement
5. Control

> **Stop and Think 17.5**
> Try to break down the above three problem-solving sequences to PDSA.
> What are the similarities? What are the differences?

17.5.3. QC circles

Quality control circles – also known as "QC circles" or "quality circles" – are perhaps the most celebrated quality improvement tool to come out of Japan in the 1960s. In the US and Europe, QC circles were briefly popular in the early 1980s but petered out within a few years, most likely for being oversold as a complete approach when, in fact, they are only one minor component of quality management. They still exist and thrive in Japan and Asia, with yearly international conventions (ICQCC) rotating between Asian capitals.

A QC circle is a group of 5 to 8 employees who meet once or twice a week, and occasionally on weekends, to solve quality problems. It's a one-year engagement for participants, and a circle may take on 3 to 4 projects during that time. The participants are mostly individual contributors, whether on the production floor or in offices, they are paid overtime and management supports the activity through training, by attending kick-off and report-out meetings, and by providing support from Engineering, Maintenance, or Quality as needed.

In a factory with an active QC circle program, about 20% of the workforce may be participating at a time, with successful projects featured at town hall meetings and award ceremonies. Participation is voluntary but management support is clear and visible, with managers walking a fine line between gently nudging operators and applying undue pressure.

QC circles serve multiple purposes:

1. They address a class of problems that would otherwise be left unsolved.
2. They leverage the creativity of rank-and-file employees while vetting their ideas through the circles.
3. They develop the technical and project management skills of the workforce.
4. They reveal the management potential of individuals who take on circle leadership.
5. They provide rich learning opportunities for everyone involved.
6. They improve morale.

The technical training provided to QC circle members is a key difference compared to earlier attempts at small group activity. The content came to be called the *7 tools of QC* (Box 16.4 on page 650), as published by Kaoru Ishikawa in 1970. Variants of these 7 tools are still being promoted by the ASQ for use outside the context of QC circles, with spreadsheet templates. The main legacy of QC circles is not the 7 tools but the value of providing targeted technical training when engaging employees in small group activity. Nothing stops a team in its tracks faster than tasking it with solving a problem that its members feel they are not competent to address.

For example, if management in a plastics extrusion plant sets the goal of reducing raw material losses at machine startup, the team will not be able to do much unless it includes a *process engineer* who understands the consequences of tweaking machine parameters and input flow. Companies that offer classes on process physics and chemistry to all employees – at the popular science level – find that employees hunger for this knowledge, and that it enhances their ability to contribute in improvement projects.

The content of project-specific training varies with the kind of projects circles are targeted for. When a team starts a SMED project, it receives training on the SMED methodology; a cell project, on cell design; kanban implementation, on the rules for using kanban. In all these cases, as for QC circles, the skills are promptly forgotten unless put to use immediately.

17.5.4. Quick Response Quality Control

Nissan's Quick Response Quality Control (QRQC) is a management approach. It's about organizing the response to quality problems, not about the technical tools used to solve them. It is intended to help detect problems, solve them, and document solutions, thereby growing the skills of the workforce. QRQC neither mandates nor excludes mistake-proofing or any statistical/data science tool. Where other methods tend to create long lists of to-dos, QRQC promotes a just-do-it mentality.

QRQC is not well documented in the English literature. It originated at Nissan and spread to France's automotive and aerospace industries through Nissan's cooperation with Renault, starting in 1999. *Perfect QRQC, Vol. 1 and 2*, in French by Hakim Aoudia and Quintin Testa[33] are rare books dedicated to QRQC. In Japanese, the Nissan Production Way (NPW) (日産生産方式, 2005) lists it among its "25 keywords" and devotes four pages to it. Nissan design engineer Kazuo Kawashima is generally credited with inventing QRQC. He transferred to production at the Nissan Kyushu plant in 1993.

At the Kyushu plant, Kawashima had observed that the solution of quality problems was delayed by the lack of peer-to-peer collaboration at the level in the organization that was closest to the point of occurrence (POO). The organization was in silos and, for example, to obtain engineering resources for a shop floor problem, a team leader's request wound its way up and down the hierarchy before it could be approved. Not only did this process lose time, but it filtered and mangled the information beyond recognition by the time anyone could act on it. His goal with QRQC was to have teams with all the requisite skills acting autonomously, as close to the POO as possible.

QRQC focuses on quick response to anomalies and defects occurring on the production shop floor. It is a generic term for a system that now targets both quality and other problems. QRQC includes both (1) action in real-time on the production line and (2) longer PDCA cycles on a shift or daily basis, involving engineers and managers:

- In **real-time QRQC**, each production operator detects anomalies and malfunctions as they occur on each workpiece. This is a mechanism to respond on the spot, in addition to the quality checkpoints. It is also a structure to prevent defectives from escaping downstream.
- **Daily or shift QRQC** is for the management system to promote fundamental problem-solving based on data for the anomalies and malfunctions uncovered by real-time QRQC.

QRQC aims to address each problem by the people closest to it, at the lowest possible level in the management hierarchy. This requires a structure and an escalation process, from the production line, to the team or shop level, and to the whole plant level. The goals of QRQC are similar to those of Toyota's *Ji Koutei Kanketsu* (JKK) translating to "building in quality with ownership:"

1. To prevent defectives from moving down the production line. Do not make defectives and, if you do, do not pass it on! Deal with it on the spot.
2. To make problems visible and solve them quickly. In principle, problems revealed by real-time QRQC are analyzed and countermeasures taken by the next day. Problems that technically require analysis are solved in a week.
3. To make each operator quality-conscious and make the entire production sensitive and responsive to quality issues.

33 Aoudia, H. & Testa, Q. (2012). *Perfect QRQC (Quick Response Quality Control)*. Maxima. / Aoudia, H. (2015). *Perfect QRQC: Volume 2, Prevention, standardization, coaching*. Maxima.

When a problem is detected, the relevant actors are called together at the place where the problem was detected or happened. The QRQC literature talks about "the three actuals," an alternative formulation of Toyota's genchi genbutsu, which means to go where it happens, focus on reality, and deal with actual objects. At the actual place, the team work through a structured list of problem-solving questions. The problem descriptions answer the classic *5W1H questions* – "who, what, where, when, why, how" – with the addition of "how many?" (Box 17.6), and the *five whys*. The system relies on forms, usually large flip charts, to ensure that the teams answer all questions (see Figure 17.22).

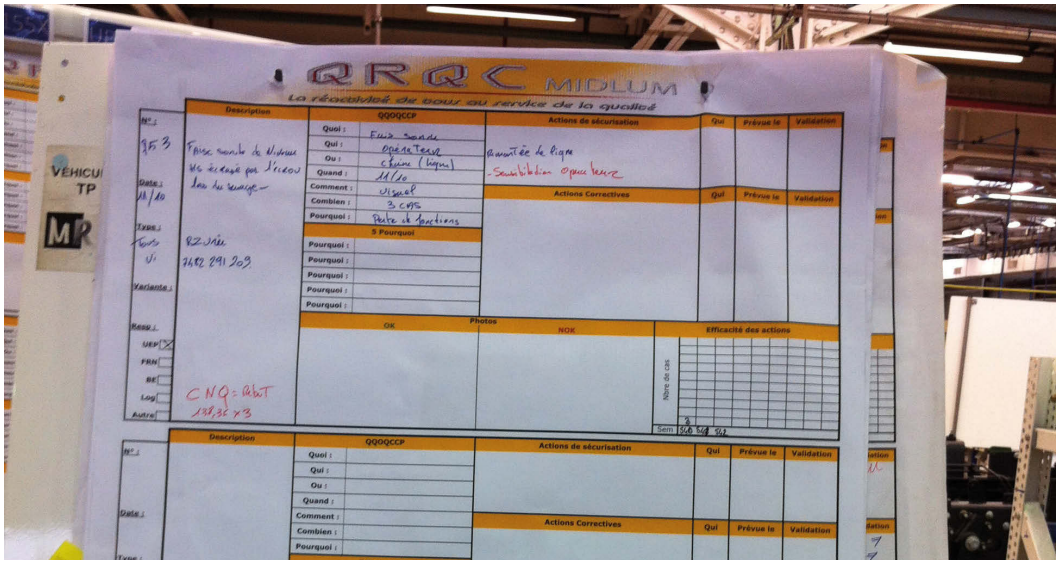


Figure 17.22 QRQC flipover templates used in Renault Trucks

Assuming that the plant practices real-time QRQC, each production operator or quality checker worker immediately informs the supervisor of any abnormality or defect and takes countermeasures on the spot. When taking countermeasures is too difficult, they make sure the next operation is notified of the need for corrective action. If this is not in place, too many items pop up at final inspection to analyze and identify trends daily. Once it is, the daily QRQC proceeds as follows:

1. **Data entry and information sharing.** Final inspection captures the abnormalities and defects that escaped the production process and the quality checks and enters them into the information system in real-time for communication to all departments.
2. **Preparation of QRQC meeting.** Participants in the QRQC meeting decide what to report at the meeting based on slicing and dicing the data, analyzing and identifying trends from all operations, and final inspection.
3. **The QRQC meeting.** Supervisors and members of the support departments meet for 30 to 40 minutes every morning to (1) share news – including items other than quality – (2) assign responsibility for countermeasures, (3) track the progress of countermeasures, and (4) confirm their effectiveness. They confirm results for yesterday's problems and issue instructions for today's. They document the decisions in forms for sharing.
4. **Analysis and implementation of countermeasures.** Following the directions of the QRQC meeting, the representatives from each department analyze the defects and devise countermeasures. For abnormalities and defects caused by product or process design rather than the production line, a lead time of a week is allowed to define the countermeasures.
5. **Ensuring durable countermeasures.** You incorporate them into the work instructions and update other master data as needed.

Box 17.6 **5W1H or "The Kipling method"**

In 1902, British journalist, poet, and novelist Rudyard Kipling wrote a poem about children's curiosity and sense of wonder. In his children's storybook *Just So Stories*,[34] one of the poems is titled *Six Honest Serving Men*. It begins as follows:

> *I keep six honest serving-men*
> *(They taught me all I knew);*
> *Their names are What and Why and When*
> *And How and Where and Who (...)*

Besides highlighting the importance of the "5W1H," Kipling makes the point that children ask many more questions than grown ups and prefer "how," "where," and – especially – "why."

17.5.5. **Root cause analysis**

Whatever problem-solving method is chosen, it is important to go beyond immediate countermeasures, identify the root cause, and eliminate it to keep the problem from recurring. Root cause analysis (RCA) is the use of structured methods to identify the underlying causes of errors.

Five whys

The simplest "tool" for getting to the root cause of an observed problem is asking "why" five times. The point of asking multiple times is to keep teams from stopping with obvious, simple answers and make them dig deeper. The number five is arbitrary and should not be taken as an order. There usually is more than one path of possible causes and, in reality, it's more like exploring a tree.

Fishbone diagram

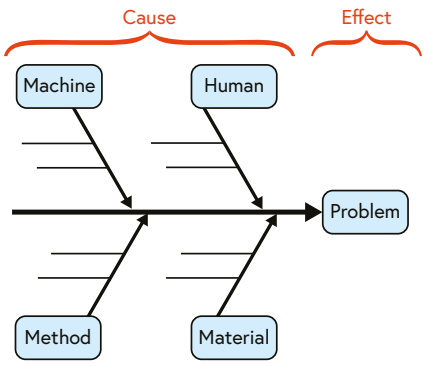


Figure 17.23 A typical fishbone diagram template

In this spirit, the fishbone diagram, or Ishikawa diagram, or cause-and-effect diagram, is a tree of possible causes. The causal relationships are not precisely defined, which is deliberate because requiring them to be as specific as in, say, a fault tree would stop brainstorming. A fishbone diagram usually breaks down an observed problem into possible problem causes, separated into human, machine, material, and method (the "4Ms"), or with additional categories like "milieu," "measurement," and "management" (Figure 17.23).

34 Kipling, R. (2009). *Just so stories for little children*. Oxford University Press. (Original work published 1902).

Attractive though it may be in a classroom, the fishbone diagram is rarely used in actual projects. In quality problem-solving, Honda, for example, uses *why-why analysis* instead, the first "why" being the reason for the occurrence of a defect, and the second the reason it was not immediately detected. You get to ask why five times in each column but the result is not formatted as a fishbone diagram.

Check sheets

One of Ishsikawa's basic tools is the "*checks sheet*." A check sheet is a data recording form that has been designed in a way that makes it intuitively easy to interpret. Check sheets are a manual means of aggregating events by category that, once filled out, look like a horizontal bar chart. Anyone can understand a check sheet but it is slow, labor-intensive, and error-prone, and it only produces a summary.

Consider a painting process. After painting, the operator inspects the results and notes down problems and frequency in the check sheet. Figure 17.24 show that inadequate coverage (17 incidents) and chip (14) are the two most frequently recorded quality error in this process, amounting to 77% of the errors in the process. It also shows how error-prone the collection method is and a clever alternative, which was proposed in Japan.

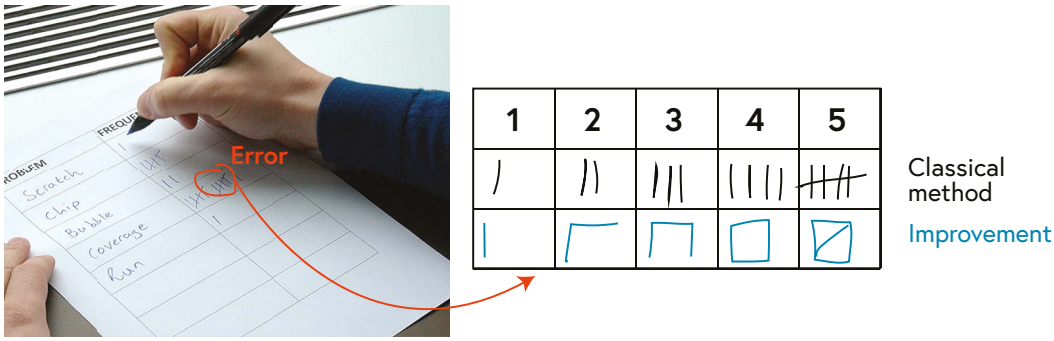


Figure 17.24 An example of a check sheet from a painting process

Check sheets are simple and everyone understands them without training. Collecting data in this fashion makes you observe the outcomes more closely than you otherwise would.

But how do you teach this method to a generation of operators who have grown up using smartphones? Being on the lookout for items to check keeps you eyes focused on the operation and the product but you really can't justify using a technique based only on its side effects. The primary purpose of data collection is to collect data. A problem with this simple recording form is that the complete event log is lost. For example, one records how many paint scratches you observed but not when or where exactly they occurred, which might provide clues as to why.

If you have an event log, with categories assigned to each event, then you can cross-tabulate it into event counts by category. It, therefore, makes sense to replace the check sheet with a form of event logger. SCADA systems and MESs do record events, uploaded from PLCs or other types of controllers. Engineers should know how to validate their accuracy and retrieve them in the form of tables. Event logs can also be derived automatically from vision systems when errors are correctly labeled.

17.6. When defects slip through

Officially, every company seeks to improve quality. No management would ever admit to not caring about it or deliberately degrading it. There are, however, examples of companies doing exactly that, through *planned obsolescence*. When defects are not deliberate, they must be fixed. If you catch defects before they leave to the market, you can fix them with *rework operations*. If they already entered the market, you may have to issue a *recall*.

17.6.1. Planned obsolescence

Planned obsolescence is the strategy of deliberately ensuring that a version of a given product will be out of date or useless within a set time period. It is a business practice engineers should oppose but marketers and finance staff sometimes prescribe for business reasons. Designing products to become unattractive through styling changes is not an assault on quality but other policies are, like designing products to break down quickly, terminating the supply of spare parts for older models, or providing system software updates that slow down aging computers.

This business strategy started with the Phoebus cartel,[35] an organization set up in Switzerland in 1924 by the world's largest lightbulb manufacturers that deliberately reduced the average life of bulbs from a technically feasible 2,500 hours to 1,000 hours, and they remained at this level until incandescent light bulbs became obsolete in the 2000s.

By the 1950s, some appliance companies had "reliability" departments engineering products to fail as soon as the warranty expired. While this was not advertized, organization like Consumer Reports in the US noticed the degradation in quality, and these policies became known through engineers who felt uncomfortable with the task. These policies were eventually ended in the US by competition from abroad.

Today, laptops, tablets, smartphones, and even watches are all computers with operating systems that suppliers frequently update, and these updates slow down the devices, putting pressure on users to buy newer models.

In addition, unlike much of the old technology, modern technology products, which are filled with batteries, microchips, and integrated circuits, have become difficult and even dangerous to attempt to fix or repair at home. Again, environmental and cost-conscious consumers have started protest movements, like the "Right to Repair" and "Repair Café" movements. The movements have an impact: In 2012, Massachusetts passed the US's first Motor Vehicle Owners' Right to Repair Act. In France, since 2021, manufacturers of electronic products must visually mark products according to a Repairability Index. In the 2020s, the European Union is expected to pass several directives that secures consumers' right to repair.

35 Krajewski, M. (2014). The great lightbulb conspiracy. *IEEE Spectrum, 51*(10).

17.6.2. **Rework operations**

Built-in-quality control methods are imperfect, which means factories engage rework and scrap defectives. Some types of rework operations are *not* self-inflicted by manufacturers. For example, after having made over 3 billion cars in over 100 years, humans have yet to find a way to paint car bodies without having to rework at least one paint job in seven. In this kind of circumstances, exhortations to do it right the first time or to stabilize the process are unhelpful.

The assembly of military electronics also involves multiple stages of inspections, testing, and rework. After functional tests, devices go through "burn in" where they are subjected to temperature cycles, vibration tests, and leak testing. As each unit sells for prices in the tens of thousands of dollars, management invests substantial efforts in reworking units that fail anywhere along that chain.

The majority of rework that goes on in assembly, however, *is* self-inflicted by the policy of assembling incomplete product units in the presence of shortages. The most common countermeasure – assembling product with missing parts – is ineffective for the following reasons:

1. **It does not get the product into the customers' hands any sooner.** It cannot be shipped until all the parts are assembled anyway, and the assembly time is a small fraction of the order fulfilment lead time for the missing parts.
2. **It inflates the work-in-process with parts that are already committed to a specific unit of finished goods and are therefore unavailable for alternative uses.** When needed, partially built units can of course be cannibalized for parts, which requires even more labor and adds complexity to production control.
3. **When missing parts finally arrive, the work of finishing assembly is done in *repair mode*, not under the controlled conditions of the assembly line.** Access to the point of assembly may require undoing some of the work previously done, without the fixtures, tools, and instructions that ensure productivity and quality on the line.

The most tempting approach is to assign the rework task to the station where the faulty assembly was made. It has the following advantages:

- The station has all the fixtures, tooling, and instructions to perform the assembly in the standard manner.
- Returning the unit to where the defect originated is an effective way to make the assembler aware of the problem and of the need to prevent it recurring.

The limitations of this approach are equally obvious:

1. Assembly stations may need to be fitted with special tools for disassembly, which assembly stations normally don't need.
2. Defective units returned from rework interfere with the normal flow of good parts. This is particularly true with moving conveyors. Operators simply have no time available for rework.

Figure 17.25 shows three alternative ways to deal with defects in an assembly line.

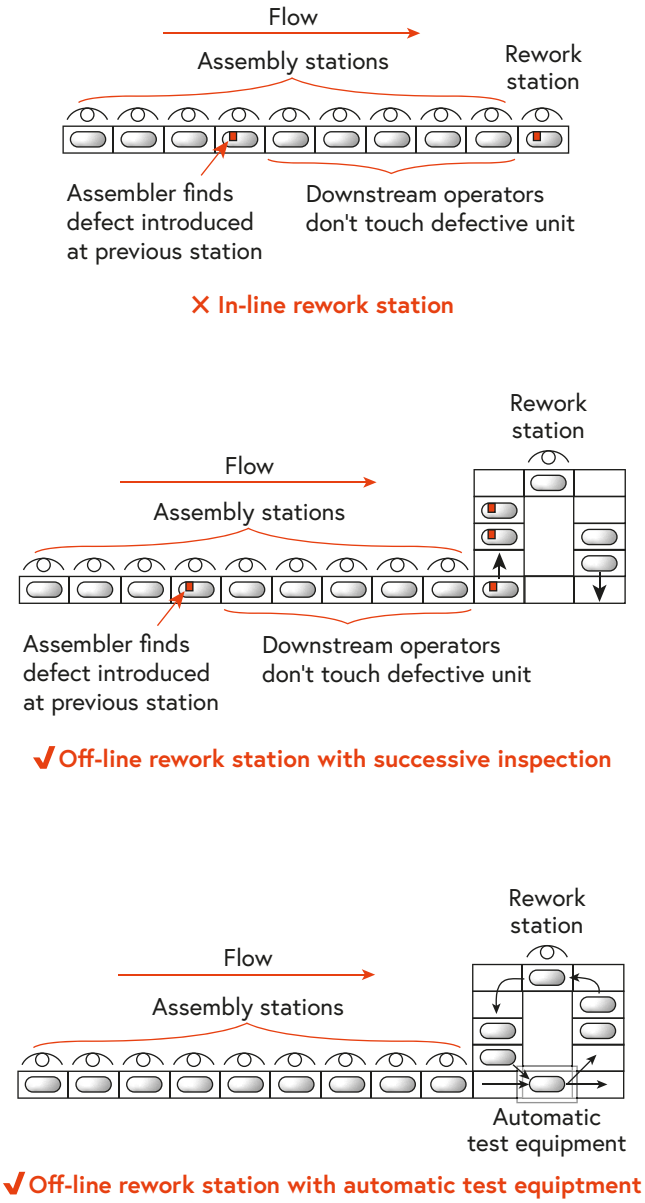


Figure 17.25 Offline rework station

Rework stations could be interspersed at multiple locations along the assembly line, as shown on the top of Figure 17.25, with the defective units tagged in between where defects are uncovered through successive inspection. This would only work, however, if the rework could always be completed within the takt time. This is unlikely, especially considering that it usually does not make sense for assemblers at the intervening stations to carry out their tasks on units that are already marked as defective. This means that the rework operator would not only have to remove the defect introduced some stations earlier but also catch up on the assembly work that should have been done at the stations in between.

The center part of Figure 17.25 shows an off-line rework station, with small buffers of defective units waiting for rework and reworked units waiting to rejoin the line. Defective units that are taken off the line create gaps on the conveyor that are filled with repaired units. This approach works in a number of circumstances, such as:

- The line makes a single product at a takt time of 1 minute.

- The rework station serves 10 assembly stations.
- Rework on these stations affects at most 10% of the units.
- The rework itself takes less than 10 minutes.

In the automotive industry, it is used in body-in-white and engine assembly. In a mixed-flow line, you can still use the approach but the rework disrupts the planned process sequence, since the defective unit that is pulled off the line is not necessarily of the same product as the reworked unit that replaces it. It is manageable as long as the sequence-change information is effectively communicated downstream, which happens if all units are self-identifying through RFID tags or other auto-ID devices.

The bottom part of Figure 17.25 shows a situation that is more typical of electronics, where the defects are found not by successive inspection but by automatic test equipment. Reworked parts return to the same tester for validation, and the tester must therefore be able to test both an normal incoming unit and a reworded unit within the takt time.

17.6.3. **Recalls**

No matter how diligently and effectively a company works to prevent defects escaping to customers, it doesn't always work, with sometimes tragic consequences, and manufacturers are obligated to recall products from the fields to replace, repair, or update them. The Quality department organizes and executes recalls, and works with public relations to answer questions from the news media. In some cases, the CEO is called upon to answer questions from the government in public hearings.

In 2000, Ford CEO Jacques Nasser faced US congress to account for the Firestone tread separation problems on tires used in the Ford Explorer; in 2010, it was Toyota CEO Akio Toyoda's turn about unwanted acceleration in some Toyota models; in 2019, it was Boeing CEO Dennis Muilenburg about the crashes of the Boeing 737 MAX. Nasser defended the Explorer and blamed the tire manufacturer; Toyoda apologized for getting Toyota's priorities of safety first confused with volume growth, in particular to victims' families, and outlined changes the company was making to prevent recurrence; Muilenburg admitted that Boeing management "made mistakes." Nasser resigned shortly afterwards, and Dennis Muilenberg was fired. Whether or not it has to do with his testimony, Akio Toyoda is still running Toyota as of 2022.

Box 17.7 presents Johnson & Johnson's celebrated recall of the Tylenol drug in 1982. A recall involves the following steps, taken under the leadership of the Quality department:

1. Notify the owners of the recalled units. If they are known, you can contact them directly. Otherwise, you have to announce the recall through ads in the media and hope to get through to all the owners, drawing attention to the recall from the public.
2. Some recalled products only need a software update that can be provided remotely. Very large products, like airliners, are repaired in the field or at maintenance shops; cars are driven by their owners to dealerships; medications are returned to pharmacies.
3. Keep the public informed of the progress of the recall.

During a recall, the company has no more urgent business than to get it done quickly and thoroughly. As a consequence, the Quality Manager has the authority to requisition resources from other departments for this purpose. At that time, also, the traceability product and process information in the history database becomes essential, when it is otherwise unused in normal operations.

Box 17.7 **Johnson & Johnson's recall of Tylenol in 1982**

In September, 1982, in the Chicago area, seven people died after ingesting cyanide-laced Tylenol, prompting a panic and a recall of the product by Johnson & Johnson. The people who poisoned the drug were never caught. Johnson & Johnson immediately lost most of its share of the pain reliever market and of the value of its stocks.

Johnson & Johnson issued a recall of unsold inventory, and offered to replace the drugs consumers had already bought with safer tablets free of charge; 31 million bottles of Tylenol capsules were recalled. Johnson & Johnson also worked with the US Food and Drug Administration (FDA) to develop tamper-proof packages (Figure 17.26). The chairman of Johnson & Johnson publicly and transparently explained to the media the chain of events.



Figure 17.26 Advertisement of tamper-resistant Tylenol medicine in 1982 (Credit: Johnson & Johnson)

Within a year, Johnson & Johnson had recovered its market share and its recall response has been cited as a model of corporate emergency response since then. The event lead to changes in federal US law. For example, it became a requirement to package all medicine with tamper-resistant technology like blisters, seals, and shrink wrap bottle covers, and all medicine got a warning label urging the consumer not to use the product if the seals were broken. The Tylenol case is a standard business case taught in risk management.

Tylenol products have ben recalled multiple times after the event in 1982. By now, recalls are standard procedures in pharmaceutical industries.

Summary

The most essential points in this chapter were as follows:

1. Quality is a fundamental capability for any manufacturer that needs to be pursued in a strategic manner.
2. Inspection and testing is an inevitable part of manufacturing involving decisions like sample testing versus inspecting every part.
3. The enemy of quality is variation.
4. Many data-driven methods are developed to find and fix quality errors:
   - Statistical Process Control
   - Design of Experiments
   - Six sigma
   - Methods based on data science.
5. One-piece flow and first-in-first-out allow for faster quality error detection and help to build quality into the processes.
6. Instead of fixing quality mistakes after they appeared, manufacturers should seek error-proof processes with poka-yoke so that errors cannot happen at all.
7. If quality errors slip through a process one needs to act on it by engaging in problem-solving activities:
   - PDSA/PDCA is a generic problem-solving cycle.
   - 8D, Imai's 7 steps, and DMAIC are sequential problem-solving methods.
   - Quick Response Quality Control (QRQC) aims to fix errors immediately after they happen at the lowest level possible.
   - Quality control circles are small-group activities that tackle quality problems on a regular basis.
   - Root cause analyses are tools for identifying the true source of a quality error and include simple tools like fishbone analysis and the five whys.
8. Quality errors also need to be fixed. If they are identified internally, they may have to be reworked. If they are identified after they are received by the customer, companies may have to issue a recall.

Test yourself

1. What is the role of the Quality department?
2. What is a control chart?
3. What is the role of data science in quality management?
4. Should operators inspect their own work? What are the pros and cons?
5. How do you setup a simple factorial experiment?

6 What is meant by built-in-quality? How can one achieve it?
7 What is the difference between poka-yoke and usability engineering?
8 What is PDSA and how does it work?
9 What is QRQC and how does it work?
10 When should you handle rework online at the station it appeared and when should you let the part pass through all stations to a rework parking lot?

Further reading

Galsworth, G. D. (2017). *Visual workplace: Visual thinking*. Productivity Press.

Garvin, D. A. (1988). *Managing quality: The strategic and competitive edge*. Free Press.

Hackman, J. R. & Wageman, R. (1995). Total Quality Management: Empirical, conceptual, and practical issues. *Administrative Science Quarterly, 40*(2), 309–342.

Hoyle, D. (2007). *Quality management essentials*. Routledge.

Ishikawa, K. (1985). *What is Total Quality Control?* Prentice Hall.

Juran, J. M. (1988). *Juran on planning for quality*. Free Press.

Norman, D. (1990). *The design of everyday things*. Doubleday.

Phadke, M. S. (1989). *Quality engineering using Design of Experiments*. In: Dehnad, K. (Ed.). Quality control, robust design, and the Taguchi Method. Springer.

Pyzdek, T. & Keller, P. (2013). *The handbook for quality management*. McGraw-Hill.

Senoner, J. (2021). *Artificial intelligence in manufacturing: Augmenting humans at work* [Doctoral Thesis, ETH Zurich].

Shah, R., Ball, G. P., & Netessine, S. (2017). Plant operations and product recalls in the automotive industry: An empirical investigation. *Management Science, 63(*8), 2439–2459.

Sousa, R. & Voss, C. A. (2002). Quality management re-visited: A reflective review and agenda for future research. *Journal of Operations Management, 20*(1), 91–109.

Taguchi, G. (1986). *Introduction to quality engineering: Designing quality into products and processes*. Asian Productivity Organization, Tokyo.

Taguchi, G., Chowdhury, S., & Wu, Y. (2004). *Taguchi's quality engineering handbook*. John Wiley & Sons, Inc.

18 Maintenance

Besides souring, installing, and programming machines, maintaining them is the task that will ultimately remain for people in the future. The term "maintenance" covers restoring equipment to service after a failure, preventing or slowing down its deterioration, and satisfying formal requirements from customers or government agencies. Together with major failures, some high-speed machines are affected by microstoppages that are quickly resolved by production operators and not reported to Maintenance. Microstoppages are individually small problems but collectively drain operator productivity.

In steel mills or chemical plants, the total number of people engaged in maintenance at one time may temporarily exceed the number of production operators. The scope of this activity in manufacturing is such that one would expect it to receive more attention than it does. Maintenance staff get all possible credit when they fix a broken machine, but nobody thanks them for the breakdowns that never happen.

This chapter introduces maintenance. It summarizes what an industrial manager or engineer needs to know about it but it is not a deep-dive intended for professional maintenance technicians. Approaches that are specific to maintenance include Reliability-Centered Maintenance (RCM) to set component replacement and redundancy policies based on Failure Mode and Effects Analysis (FMEA), and Total Productive Maintenance (TPM) to involve all employees in maintaining equipment and organize the involvement of specialists.

Learning objectives

After working through this chapter, you should be able to do the following:

1. Explain the role of maintenance in manufacturing.
2. Make a maintenance strategy.
3. Staff and organize a maintenance department.
4. Use 5S to organize and maintain a workspace.
5. Use Failure Mode and Effects Analysis (FMEA).
6. Explain Reliability-Centered Maintenance (RCM).
7. Explain Total Productive Maintenance (TPM).
8. Explain predictive maintenance and prescriptive maintenance.

DOI:10.4324/9781351110310-23

18.1. Introduction to maintenance

> **Definition**:
> Maintenance is the task of preserving a system's functional state.

Maintenance encompasses all the work needed in a factory to keep it running. It produces no output visible to customers and, for this reason, is often undervalued and targeted in cost-cutting. When done without any engineering analysis of the needs, postponing or canceling maintenance activities, while producing short-term savings, eventually increases maintenance costs.

The objects of maintenance include tools, jigs, and fixtures to machines, transportation equipment, production lines, and entire facilities. In manufacturing, the term "maintenance" is usually about hardware. It is also used about software, with a different meaning, to describe the development and deployment of upgrades.

Parts can be *repaired* or *replaced*. In advanced economies, repairs consist almost entirely of component replacements. If an electric motor breaks down, you replace it with a new one. In emerging economies, manufacturers often rely on old machines for which spare parts may not be available. When an electric motor breaks down, they rewind it.

18.1.1. Preventive versus corrective maintenance

There are different types of maintenance work. It can generally be split into *preventive* versus *corrective* activities, which broadly overlap with *planned* versus *unplanned* maintenance activities. The activities vary in terms of urgency, complexity, and resource needs.

Planned activities are always preventive because they seek to prevent failures from happening. Planned activities can be split into three activities:

1. Daily housekeeping:
   - Removing unnecessary items from workspace
   - Assigning locations to all necessary items and making sure that they are where they should be
   - Cleaning machines and workstations
2. Small maintenance:
   - Replacing frayed cables
   - Plugging leaks
   - Replacing bulbs
   - Cleaning filters
3. Periodic maintenance:
   - Monthly, quarterly, yearly
   - Usage-based
   - Overhauls
   - Upgrades

Unplanned activities are usually corrective and involve repairing *after* failures. However, if the factory can predict failures and issue alarms, unplanned activities can also be preventive in that they take place

before a malfunction causes a failure. This is *predictive maintenance*. Unplanned activities can be split into three activities:

1. Resolving microstoppages
2. Minor repairs
3. Major repairs

18.1.2. Maintenance strategy

A Maintenance department should decide a *maintenance strategy*. The strategy can utilize a mix of preventive and corrective actions. Figure 18.1 summarizes the overall options.

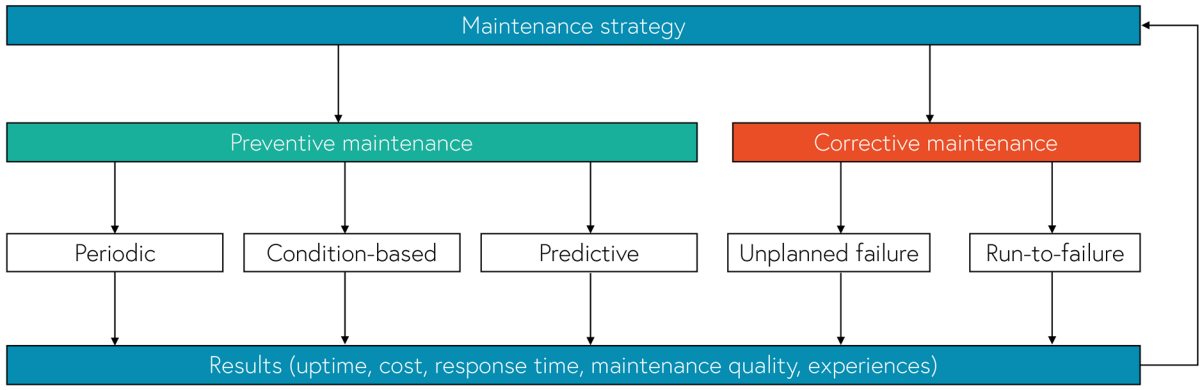


Figure 18.1 Maintenance strategies

Preventive maintenance can be *periodic,* as in changing the battery of the fire alarm every December 1; *condition-based*, as in the fire alarm beeping before the battery is empty or checking it during planned safety inspections; or *predictive*, as in calculating its remaining useful life with analytics. Corrective maintenance is needed when equipment is *run until it fails*, in which case the failure is expected but its timing cannot be foreseen, and when equipment has *unplanned failures*. Having a strategy means considering what equipment should be targeted for the different types of maintenance, a plan for how to prioritize and execute maintenance tasks, and a feedback process where Maintenance learns from its own performance and ultimate maintenance results.

Logically, most maintenance activities should be allocated to preventive maintenance, with slack time for corrective maintenance. Maintaining a machine according to a plan is easier than fixing it when it fails. Even if the duration of a planned activity depends on what the technicians find when they open the machine, it is still less random.

Unfortunately, in actual operations, it is often the opposite. Maintenance technicians are so busy with corrective maintenance that they don't find time for preventive maintenance, and the performance of the factory drops. Mapping results and actual activities to the maintenance strategy help bring it back on track. Responding to failures always takes priority over planned activities "that can be postponed" but repeated postponements end up causing more failures and more disruptions of the plans. The key to avoiding this vicious cycle is to give the Maintenance department just enough slack capacity that there is almost always one technician available to respond.

To support the strategy, the Maintenance department must also administer the process, keeping records, monitoring equipment, informing platforms users, managing spare parts logistics, and training staff. Having extra technicians with the required skills is one way to provide that extra capacity but another one is improving the planned maintenance.

> **Stop and Think 18.1**
> In your daily life, can you give some examples of preventive and corrective maintenance?

18.2. **Organizing for maintenance**

Maintenance has traditionally been organized by *trade*s, with titles like millwright, pipe fitter, mechanic, welder, or electrician, as illustrated in Figure 18.2, with almost every maintenance job requiring contributions from multiple trades. To these categories, the advent of OT has added electronic controls.

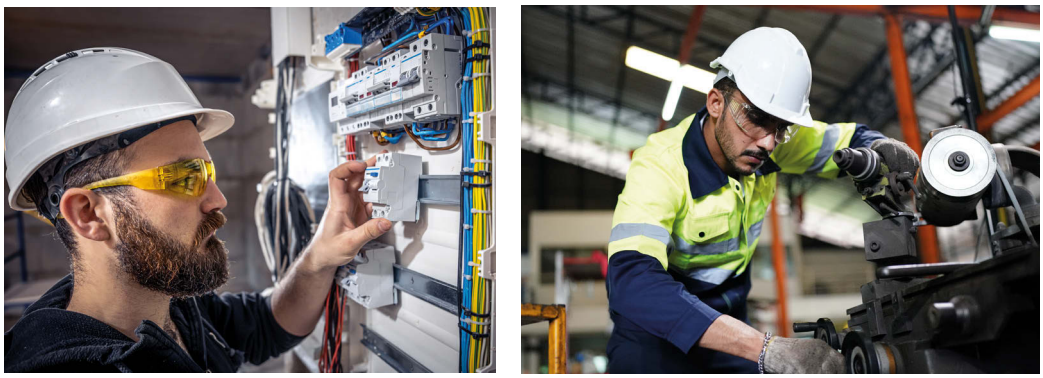


Figure 18.2 Electrician and millwright

The challenge is to organize maintenance to reap the benefits or special skills while delegating simple repairs to handypersons, without requiring the coordination of three or four skilled trades.

In factories, multiple actors participate in maintenance:

- **Maintenance technicians**, accounting for a small fraction of the factory's total headcount, check machines and replace components and consumables like filters according to plan, and repair machines as needed.
- **Production operators** keep workstations tidy, visible, and clean through 5S activity, resolve microstoppages, and report any anomalies they encounter to the Maintenance department.
- **Service technicians and engineers**, dispatched by equipment suppliers, assist in-house technicians and learn the specific ways this factory uses their equipment.
- **Outside contractors** help during periodic shutdowns and overhaul equipment off-site.

Support functions are also directly or indirectly involved in maintenance:

- **Production planners** schedule planned maintenance in coordination with Maintenance.

- **Human resources** hire and organize the training of maintenance staff in coordination with Maintenance.
- **Purchasing** negotiates supplier contracts and buys machines and parts.
- **Quality** and the improvement office draw on maintenance staff for help in problem solving.

18.2.1. Maintenance staff

Maintenance technicians are typically either "generalists" or "specialists." Often drawn from the ranks of experienced operators, the "*generalists*" are on call for troubleshooting in a geographical section of the shop, preferably one with equipment they are familiar with from having operated it. They are trained to safely replace dirty filters, frayed cables, and leaky hoses, and to shut down and restart electronic controls. They know which materials are hazardous and how to protect themselves. As their area of responsibility includes multiple machines with different functions, they usually have a qualitative understanding of what the machines do but not a quantitative understanding of process physics. While they are assigned to a section, they also help as needed in others. If busy when a problem occurs, a local technician can call on colleagues for help but remains responsible for the outcome.

As in healthcare, the *specialists* back up the generalists when they are stumped. The specialists may be traditional millwrights, pipefitters, or electricians, or they may be expert in computer controls. Sometimes, solving problems also requires process or manufacturing engineers.



Figure 18.3 Maintenance staff at work

In healthcare, patients do not visit the doctor to get a Band-Aid. Likewise, operators do not call Maintenance if a problem is small enough. Operators resolve *microstoppages*. When production operators clear a jam and restart the machine, they are taking on what is, strictly speaking, a maintenance task but it happens too often and is too simple to justify calling Maintenance. In addition, it makes sense to delegate to operators *routine activities* that are strictly speaking *housekeeping* but help maintenance by enhancing visibility, facilitating access, and ruling out what would otherwise be possible causes of problems. This helps focus maintenance technicians on actions that require their skills.

18.2.2. The Maintenance department

A small company using 25 machine tools to make aircraft fittings may only have one engineer doing everything from process planning to maintenance. A car plant assembling 1,000 vehicles a day may have 300 people focused on just maintenance. This raises the question of right-sizing the Maintenance

department for a given factory. A related question is where to locate the staff; should they all belong to a central Maintenance department, or should they be distributed to local areas?

What is the right size of the Maintenance department?

Since manufacturing machinery varies in complexity and reliability, no single formula can provide the right Maintenance department size and location. The variety of machinery matters, and so do skills. In maintenance, it is not unusual for factories to have one or a few experienced "MacGyvers" who can fix almost any kind of machine and have others that are trainees. Another factor that affects the right-sizing of maintenance staff is the skills, dedication, and initiative of the production workforce. When machines are well understood and cared for by their users, the need for maintenance is reduced. Moreover, new machines may be so specialized that they can only be maintained by the company that produces them, increasing a factory's reliance on external service technicians.

So what is the right size of the Maintenance department? It depends on the set of factors discussed above. Maintenance's share of a plant's workforce varies by industry. In labor-intensive manufacturing activities, maintenance's share of the labor force is on the order of 5%, and it is increasing due to increased automation. For example, an auto plant that employs 5,000 today may produce the same amount with the same depth of manufacturing with 500 people 25 years from now — if cars are still made… And, out of these 500 people, 150 to 200 will be in Maintenance, the rest being primarily programmers of automatic machines. In steelworks, on the other hand, it can be 30%. Steel is an industry that has had enormous productivity increases in the past decades, with a typical plant going from 45,000 employees in the 1940s to 2,100 today for the same output.

In principle, high utilization and high availability of technicians are contradictory pursuits (recall queuing theory and Kingman's formula in Box 3.4 on page 94). If all technicians are always busy, then none is ever available, and vice versa. A sufficiently large *team* of technicians with interchangeable or just overlapping skills, however, can combine high utilization and high availability.

The Maintenance department is most effective when it does not have to be called to fix problems because their planned maintenance routines prevent failures. In such situations, senior managers may overlook the value of maintenance, and target the department for downsizing. This may shift its focus from prevention to reaction. Maintenance will be "busier" but less effective. Soon productivity will go down, and more maintenance staff is likely to be hired again.

Where to organize the maintenance personnel?

For the question of where to locate the Maintenance department – centrally in an all-factory unit or locally in dispersed subunits teams – in particular the size of the factory and the variety of technologies deployed must be taken into account. Box 18.1 gives one simplistic decision-support model.

Box 18.1 **Central maintenance group versus local maintenance teams?**

A car parts factory has a maintenance department with 20 members and considers splitting it into 4 groups of 5 members serving four areas of the factory. A car assembly factory has 300 technicians serving the entire plant and considers dividing them into 6 groups of 50. The managers of both plans want the technicians in close contact and constant communication with the production people they support. They don't want the technicians to be strangers dispatched by a distant bureaucracy. Should the central groups split up?

A simplistic model is to compare maintenance staff availability. Let us assume the technicians work in parallel, have approximately the same skills, and are busy 80% of the time when combining planned and unplanned tasks. Then we get as follows:

1. Car part factory:
   - Local teams of 5 technicians will all be busy simultaneously 0.80^5 = 33% of the time. As a consequence, one equipment failure in three has no technician available to respond. It is not an acceptable level of service.
   - In a central group of 20 technicians, all will be busy simultaneously 0.80^{20} = 1% of the time. It means that there is at least one technician available to respond immediately to 99% of the failures, unquestionably a better service.
   - In this case, the central group is better.
2. Car assembly factory:
   - Local groups of 50 technicians will all be simultaneously busy 0.8^{50} = 14 ppm of the time. This is negligible. Local groups can be used in this case.

The responsiveness discussion above is about *first responders* assumed to have the same skills. Not all maintenance work is responding to emergencies and much maintenance work requires more than one technician. It is a combination of planned and unplanned work but even planned maintenance on a machine takes more or less time depending on what the technicians discover when they open it up. You know when a yearly check on a machine is planned to start, but not whether it will take a day or a week. Skills vary widely.

While this model is simplistic, it suggests a rule-of-thumb for whether or not to split Maintenance into groups. If we call the probability of having at least one technician available when a machine breaks down "Responsiveness," then the number of technicians needed to achieve a given responsiveness while keeping the technician utilization at a given level is:

$$Number\ of\ technicians = \frac{log(1 - Responsiveness)}{log(Utilization)}$$

The flaw in the math is that when all technicians are busy, it means that there are at least as many machines needing attention as there are technicians, and not *exactly* as many. In addition, there is more to solving an equipment problem than dispatching one individual technician.

Training for maintenance

Maintenance technicians were traditionally trained to hear small noises and detect the slightest whiff of burnt rubber out of a machine. Today, they may have to decipher the meaning of a UNIX error message on the screen of a machine controller. The spread of computers and electronics in the workshops creates new needs for maintenance know-how, and therefore new training needs.

18.2.3. Planning preventive maintenance activities

Studies show that "wrench time" for maintenance staff is around 25–35%. Work time is frittered away walking and searching for parts, people, and tools. Better planning can boost the productive time to above 50%.

The key characteristic of *planned* activities is that, theoretically, they can be planned. In reality, the time needed for a planned maintenance task actually depends on what the technicians find when they open the equipment. The only time that can be known is how long they would need if they found nothing wrong. A checkup that would never find anything wrong, however, would be a waste of time. These maintenance tasks exist because, as machines get used or age, columns randomly crack, bolts loosen, pipes leak, cables fray, and so on. As a consequence, both scheduling and capacity planning for these activities must account for randomness in their durations.

The cost of preventive versus corrective maintenance is not the same in all industries. For example, when the needle breaks in a sewing machine, it is cheap to replace it and the problems caused are limited. While it may not be the best strategy, a corrective maintenance strategy for sewing machines in addition to some weekly oiling and cleaning activities can be sufficient. Quite on the opposite side, when a nuclear reactor melts down, it is a catastrophe. The more severe the effect of a failure, the more important it is to plan preventive maintenance. Box 18.2 elaborates on maintenance in nuclear power plants, and Box 18.3 compares the maintenance performance of the Japanese and French nuclear industries in the 1980s.

18.2.4. Maintenance outsourcing

In most organizations, the scope of maintenance exceeds the capacity or the capabilities of in-house resources and requires the temporary engagement of outsiders:

- **Yearly shutdowns.** Bringing in outside contractors during the yearly shutdown of a chemical plant increases the amount of maintenance work doable during a given time or shortens the time needed to perform required tasks.
- **Equipment overhauls or refurbishments.** These actions are aimed at making used equipment perform like new and commonly require shipping it back to the supplier. Specialized factories disassemble the equipment, replace its worn components, and reassemble it.
- **Special skills.** When equipment is new, the supplier's technicians and engineers know it better than the users and provide training and support to in-house maintenance. Once a factory has used it for years and modified it for its own purposes, however, the tables are turned. In-house users and maintenance technicians know it better than the supplier, for their specific use.
- **Remote maintenance.** The supplier monitors the equipment online, ships replacement parts as needed and prompts the local technicians to install them, or dispatches its own technicians to help.

Box 18.2 **Maintenance in nuclear power plants**

In maintenance, manufacturing can learn from other industries, like aviation and nuclear power, where effective maintenance is essential to safe operation. In a nuclear power plant, the maintenance program covers all structures, systems, or components within the site.

Most of the *reactors*' planned maintenance is organized around refueling outages. About once a year, each reactor shuts down for about several weeks to replace the spent fuel rods with new ones, and hundreds of maintenance technicians swarm over the reactor building to maintain devices that are inaccessible while the reactor is running. These operations involve lowering and raising the water level inside the reactor multiple times, exposing and submerging different devices. Hence, the maintenance work is organized around accessibility.

Different approaches have been used in different countries. The Japanese nuclear power industry uses many different plant designs, and has borrowed the Total Productive Maintenance (TPM) approach from manufacturing. The French industry, on the other hand, has standardized plant designs, and borrowed Reliability-Centered Maintenance (RCM) from aviation (see details in Box 18.3).



Figure 18.4 Nuclear power plant in the Netherlands

While the most routine maintenance is done in-house and equipment overhauls are outsourced, the boundary between the types of tasks that are best done either way is fuzzy. Some companies prefer to save on the permanent cost of employees, at the risk of relying on technicians who may not fully understand their application of the equipment. Others prefer to develop and retain equipment knowledge inside.

Stop and Think 18.2
Why are some industries better at maintenance than others?

Box 18.3 **Japanese versus French maintenance performance in the 1980s**

Jean-Pierre Mercier, from the French electrical utility EDF, published the chart in Figure 18.5 to compare the evolution of nuclear reactor downtime in Japan and France.[1] Each node on each of the orbits indicates performance in a given year, in terms of planned unavailability on the y-axis, and unplanned unavailability on the x-axis, and the sum of the two gives the total unavailability, so that diagonal lines indicate a constant total value.

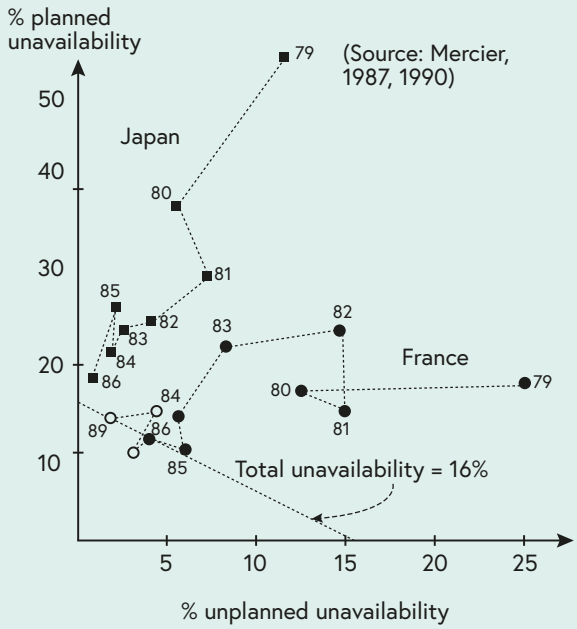


Figure 18.5 Japanese and French approaches to nuclear plant maintenance

The orbits enable us to track year-by-year progress, and improvement is marked on the chart by movement towards the origin. The Japanese and the French orbits both show improvement over the years but are so different that they do not even intersect, which begs the question of why. What did the Japanese and French industries do so differently that it produced such radically different results? Once the chart prompts you to ask the question, it is easily answered:

- France had one national utility company, with reactors of just two designs all made by the same supplier. This company reduced downtime by redesigning components or subsystems that failed in one reactor, and retrofitting the changes to all sites of the same design.
- Japan had nine private utilities and reactors with many different designs, which made the French approach impossible. Instead, they overdid preventive maintenance in the beginning, and gradually improved it, eventually achieving almost the same performance as the French utility.

1 Mercier, J.-P. (1987). *La Maintenance des Centrales Nucléaires*. Kirk I.

18.2.5. External repair centres

Diagnosis and repair methods have changed as a result of the advent of computer controls, and so has the skillset of maintenance technicians. Technicians used to rely on their senses for clues and on their experience to map clues to causes. It required them to be physically next to the machines. Today, machine controllers issue messages describing malfunctions, and run tests to identify faulty components or subsystems. For this, the maintenance technician doesn't need to be at the machine or even on-site. This process is often run from a *service center* operated by the machine supplier.

A service center performing remote diagnosis may also hold an inventory of spare components for multiple customers, with a guaranteed delivery lead time. If the faulty component requires it, a technician may travel with it to install it. If, however, the machine is designed so that it is, essentially, plug-and-play, then the local technician or even a production operator can install it, with direction as needed from the remote service center.

18.2.6. Computerized Maintenance Management Systems

The software industry has been marketing systems to support maintenance operations since the 1980s. They are known as Computerized Maintenance Management Systems (CMMS), now also known as *Enterprise Asset Management* (EAM) systems.

> **Definition**:
> CMMS is software that centralizes maintenance information and is used to administer maintenance operations.

The CMMS were originally focused on *workorder* administration. A workorder is a unit of maintenance work. Whether planned or unplanned, each intervention is defined, scheduled, and tracked as a *workorder*. It specifies the equipment, the task, and the team responsible for carrying it out. Once completed, it becomes a record of work performed, resources consumed, and technician time used.

When recorded, technical information about the task is usually entered as a *comment* by the technicians. The workorder may state that a pump was replaced and include model and serial numbers for the old and the new pumps. Any explanation of what was wrong with the old pump and how the technicians diagnosed it will be whatever they write in free text. As a consequence, *text mining* technology is required to extract information like diagnostic decision trees from these comments. Although technically possible, these technologies have yet to be taken advantage of, and, as a consequence, much of the maintenance data is never reused. For the same reason, SCADA systems are often a richer source of technical information than CMMS.

Today, CMMS are common in Maintenance departments and there are many vendors, including Hippo, SAP CMMS, Fiix, UpKeep, Fracttal, MaintainX, eMaint, and many, many others. The CMMS typically covers the following tasks:

- Workorder management
- Asset management
- Task scheduling
- Asset history recording
- External workorder generation

- Spare parts inventory management
- Audit and certification of machines
- Performance reporting

18.2.7. Spare parts management

Spare parts management is essential to maintenance. Missing spare parts are usually the main reason for long repair times. As production machinery may remain in service for decades after the supplier stops making it, many of the spare parts needed are for obsolete models, with order lead times that may be in months if the supplier has not kept them in stock.

Planning for spare parts supply is a matter of risk management, considering both the economic consequences of not having a part when needed and the cost of keeping it in stock when you don't. Hence, the goal of spare parts management is striking a balance between carrying costs and spare parts availability, performance, and reliability. Spare parts can be anything from a bolt or ball bearing to an electronic board or a pump (Figure 18.6). They usually differ for every different type of machinery and equipment, and may be needed daily or once every 10 years. Given this variety, there is no one-size-fits-all approach for spare parts inventory management.

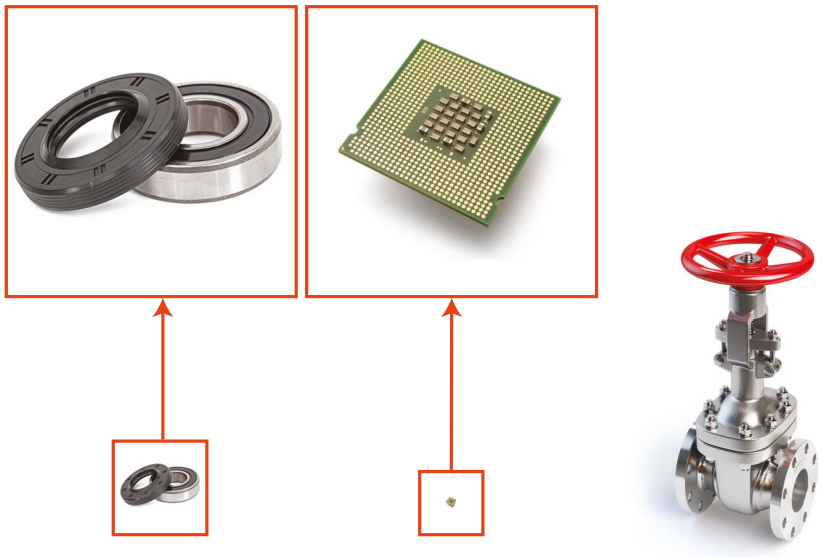


Figure 18.6 Spare parts in different criticalities, shapes, sizes, and frequency of use

Several different approaches are used in combination for different parts:

1. Keep spare parts in stock at the plant
2. Share spare parts inventory across plants
3. Buy spare parts when needed from a distributor
4. Buy spare parts when needed from the equipment manufacturer
5. Make you own spare parts with conventional or additive manufacturing (3D printing)
6. Use third-party after-market parts
7. Retrofit other parts
8. Lease/rent equipment and let the equipment supplier handle spare parts.

To decide what approach to use for different equipment and specific parts, we can use Failure Mode and Effects Analysis (FMEA) and rank parts according to the consequences of shortages. The most system-critical components we should consider holding on stock or make contractual agreements of availability with suppliers. If we hold stock, we must care for it so parts do not rust, jam, or deteriorate for other reasons. By sticking to a standard set of equipment, we can reduce the variety of spare parts.

A CMMS commonly helps you track spare parts inventory, identify parts that are no longer required, monitor spare parts use and stock levels, and decide when to restock items.

18.2.8. Maintenance role in equipment selection

In many factories, management buys equipment without consulting Maintenance, resulting in technicians crawling under tanks of hazardous materials to reach valves or sawing through a protective barrier to access a pump and welding it back together when done. In such cases, Maintenance was not sufficiently involved in the design of the installations but the opposite excess is also possible, as when the responsibility for machine selection in all plants is given to a centralized plant Maintenance department for the entire company.

In each plant, the process engineers then complain that this central department selects machines that do not facilitate the flow of production. The maintenance staff must be involved in the choice of machines so that they can do their jobs safely and efficiently, but it is not up to them to determine the operating principles of the machines in the context of a production line.

18.3. Approaches to maintenance improvement

The focus of improvement in maintenance should not be on structures and tools but on purpose. We maintain production facilities not just to comply with a mandate or fulfill formal requirements but because it allows us to deliver goods to customers without large stocks. You might add that, if your products are custom, or even if you just have high variety, there is no way you can hold stocks large enough to deliver promptly.

18.3.1. 5S workplace organization

5S is a *maintenance* approach because it keeps the workplace and factory in a neat, safe, and orderly state. It engages every employee, which makes it a stepping stone to more ambitious approaches that require participation by everyone, like TPM, described below.

"5S" refers to five Japanese words that begin with "S." American consultants have undertaken to translate these words into five English words also starting with "S" but matching the meaning is more difficult than matching the initials. The most commonly used "translations" are "Sort, Set in order, Shine, Standardize, and Sustain," but they do not fully match the meaning of the original 5S:

- Seiri (整理). In everyday Japanese, it means resolving issues like administrative problems. In 5S, it means removing from the shop floor the items that are not used routinely.

- **Seiton** (整頓). In everyday Japanese, it means arranging neatly; in 5S, having assigned locations and labels for everything you retain on the shop floor and making sure that, when not in use, each item is where it's assigned to be. A common related slogan is "a place for everything and everything in its place." What Seiton adds is that the place must be labeled.
- **Seiso** (清掃) means cleaning. Production operators clean their own workplaces at shift end, so that they notice details like spills, frayed cables, or broken lamps.
- **Seiketsu** (清潔) In everyday Japanese, it means cleanliness; in 5S, the reduction of the first three Ss to daily practice by *management enforcement*, through tools like checklists, assignment of responsibility for daily housekeeping activities, and routine audits.
- **Shitsuke** (躾). In everyday Japanese, it means upbringing and is usually used in the context of raising children or pets. It is not an action but the condition you reach when the performance of the first three Ss has become second nature to the organization. As long as you tell kids to brush their teeth every day, you are practicing *seiketsu*; once they do it without prompting, you have achieved *shitsuke*.

5S helps take care of all the small things that otherwise go unnoticed and that eventually become time thieves, sources of quality defects, or safety risks. Figure 18.7 shows two examples of the maintenance problems addressed by 5S.

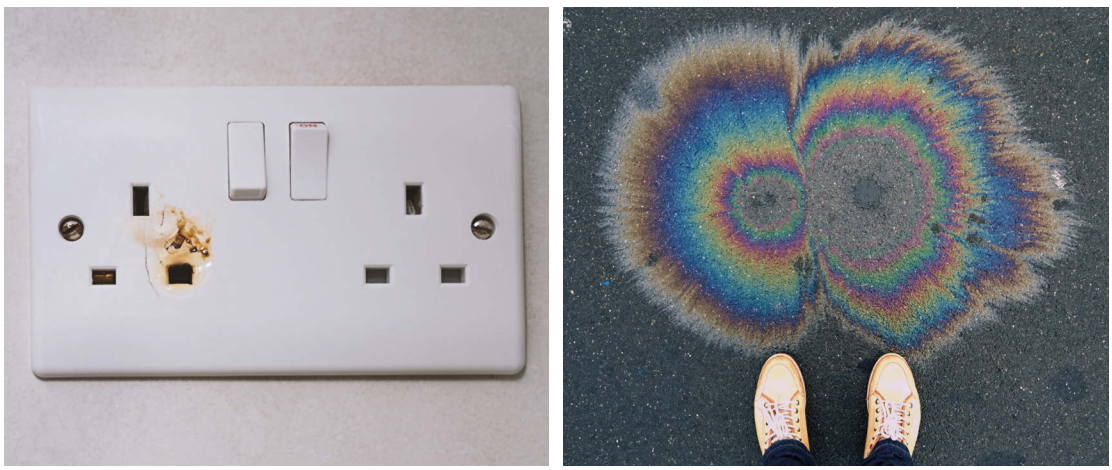


Figure 18.7 5S draws attention to small maintenance issues

Three good examples of how excellent seiri-seiton can look like beyond manufacturing are the carpenter's tool belt, the organization of a dental chair, and the chef's "mise en place," a French term for organizing ingredients and utensils for quick access before cooking.

What is essential about the third S, "cleaning" or "shining," is not what the task is but *who does it*. A janitor will wipe the oil off the floor and take it no further. If the operator does the cleaning, then the hand guides the eyes and draws attention to details that weren't there before. It works as an early warning system, and a stepping stone towards *Autonomous Maintenance*.

5S implementation challenges

5S is extremely popular and often promoted as the first "tool" to deploy in a "lean journey." The evidence is overwhelming, however, that 5S is more difficult than it seems and that starting with it is unwise.

A challenge in organizing for operators to do 5S is that it is not direct production work. Much of what we do in designing operator jobs is making sure that they are relieved of all tasks that do not directly move the product towards completion. That is why, for example, assemblers should not have to unpack parts but instead should have parts unpacked by others and presented to them within arm's reach, oriented for ease of assembly. With this logic, a production manager in a car assembly plant responded to the idea of setting aside the last 5 minutes of each shift for 5S by saying "that would cost us three cars." In reality, there is no reported case of production performance going down as a result of doing 5S, but it is not obvious.

Many managers assume 5S to be easy because it is technically trivial. It's all about human behavior, but this is exactly why it is difficult. Telling people to start by tidying up their space may evoke the same response from machinists as with teenagers. Telling a machinist in a job-shop to label hand-tool locations does not go over well either. He or she has spent the last 15 years becoming *indispensable* on specific machines. Making tools easier to find would just make it easier for *somebody else* to take over the job.

On the other hand, assume the machinist works in a cell, running multiple, different machines and rotating positions with other machinists. Then putting away each tool always in the same, clearly marked location makes sense. The machinist is valued for flexibility, not for being the only one who can operate a specific machine. Therefore, the right time for 5S is once you have changed the mode of operation.

Furthermore, 5S only works if *all employees* do it. It's a company-wide activity that involves everyone. Rather than leaping from 0% to 100% involvement, it is easier to start with focused, technical projects that only require small teams, and then leverage the success of these projects to engage everyone else. In addition, not all organizations have a use for 5S.

> **Stop and Think 18.3**
> Why do you think 5S is such an illusive improvement technique?

See Box 18.4 for the tale of two high-technology companies, Google and Intel, one of which had no use for 5S while the other did.

Box 18.4 **5S at Google versus Intel**

In *How Google Works*,[2] on p. 38, executives Eric Schmidt and Jonathan Rosenberg wrote:

> *"When offices get crowded, they tend to get messy too. Let them. When Eric [Schmidt] first arrived at Google in 2001, he asked the head of facilities, George Salah, to clean up the place. George did, and was rewarded with a note the next day from Larry Page, saying, 'Where did all my stuff go?' That random collection of stuff was an icon of a busy, stimulated workforce. […] It's OK to let your office be one hot mess."*

So the company whose mission is "to organize the world's information and make it universally accessible and useful" has no use for 5S in its offices. The explanation they give is that Google employees are "smart creatives" who do their best work in a messy environment, like Pablo Picasso in his studio. But there is another possible reason: The information that matters to the googlers' work is behind their screens, not on their desks. It's in Google's data centers, and they work on it with Google's software.

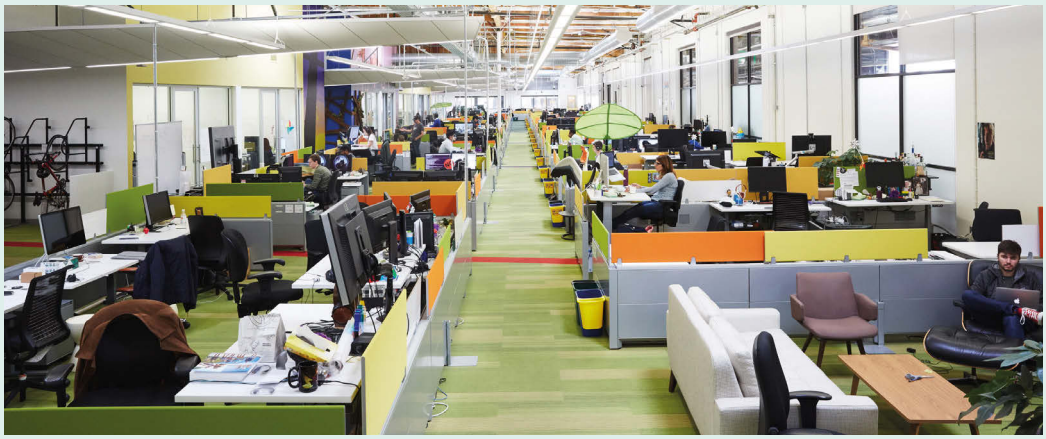


Figure 18.8 Google workspace (Credit: Google)

Not every Silicon Valley company celebrates messy desks. Intel, for one, requires employees, before leaving, to lock away all documents, clear their desktops, log out of their systems, and erase all white boards. This is to make sure that no competitor's spy, posing as a janitor, can abscond with intellectual property, and it is rooted in founder Andy Grove's conviction that "*only the paranoid survive*."[3]

As part of this "Mr. Clean" program, every facility is routinely audited, and violators written up. This is how it is explained in a handbook for contractors:

> *'Mr. Clean' is a team of two people sometimes consisting of a Senior Manager and SMS/GSS representative who inspect our buildings once a month to ensure that all facilities meet the Intel standards for cleanliness and safety. These inspections are for your benefit to aid in improving working conditions and keeping watch on safety hazards..."*

2 Rosenberg, J. & Schmidt, E. (2017). *How Google works*. Grand Central Publishing.

3 Grove, A. S. (2010). *Only the paranoid survive: How to exploit the crisis points that challenge every company*. Crown.

18.3.2. **Total Productive Maintenance**

Total Productive Maintenance (TPM) was coined in 1971 by a group led by Seiichi Nakajima at JMA in Japan, a consulting firm established in 1942 with about 1,300 employees that wanted to offer services in maintenance. The plant Maintenance department of the JMA became the Japan Institute of Plant Maintenance (JIPM), which remains the primary source of information about TPM.

TPM usually stands for "Total Productive Maintenance" but occasionally also "Total Preventive Maintenance" or "Total Productivity Maintenance." None is self-explanatory. Contrary to what a layperson might expect, "Productive Maintenance," is not about the productivity of maintenance technicians but about the effectiveness of the maintenance service to production. It is about ensuring that the equipment works when needed at its nominal rate and does not produce defectives.

As TPM became popular in *process industries*, its scope expanded beyond maintenance, and the name became the umbrella term for all improvement activities. In food or chemical plants, a layout change in packaging lines may be a "TPM" project, even though it affects much more than maintenance. The TPM proponents have suggested different versions of a TPM temple with seven or more pillars. Figure 18.9 shows a typical TPM temple.

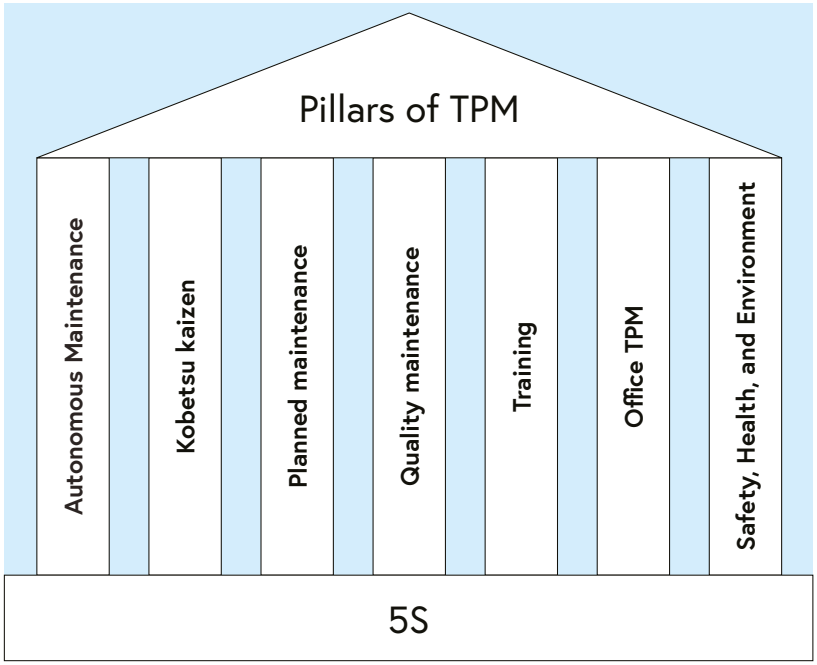


Figure 18.9 Typical representation of TPM pillars as a temple

Within TPM, *Autonomous Maintenance* is the most practical pillar. Autonomous Maintenance is a method that gives operators the responsibility for basic maintenance tasks, rather than relying on dedicated maintenance technicians. In practice, this means that tasks like safety checks, cleaning, filter replacements, and lubrication are performed by machine operators. Maintenance technicians can then concentrate on more complex maintenance tasks. Autonomous Maintenance requires light upskilling and additional responsibility and authority, which pays off in lower maintenance costs and more knowledgeable and aware operators who increase their ownership of the equipment. It also enables them to do troubleshooting and propose improvements.

18.3.3. **Reliability-Centered Maintenance**

In some cases, the goal of maintenance is to make interruptions of service less frequent and shorter. This is exactly what United Airlines pursued in the late 1960s when the Boeing 747 was introduced, and they called it *Reliability-Centered Maintenance* (RCM). Reliability is the ability of a machine or component to do what it's meant to do with no failures over time. It is different from availability, which is the probability that a machine or component is operational and accessible when needed. A telephone switch that runs 20 years without any breakdown is reliable. A phone network that lets you make calls instantly at any time is available.

RCM and the bathtub curve fallacy

Introductions to reliability theory describe the useful life of a device using "the bathtub curve" shown in Figure 18.10. According to the bathtub curve theory, a device is subject to "infant mortality" when new (Phase 1), wear-out when old (Phase 3), and have a "useful life" phase in-between with a low, constant failure rate (Phase 2). It was observed on vacuum tubes in the 1950s and assumed to apply to everything, with consequences on maintenance and part replacement policies. Obviously, this suggests a straightforward maintenance policies: You would want to monitor parts closely when new and replace them just before wear-out kicks-in. The problem is that this theory generalization was hasty.

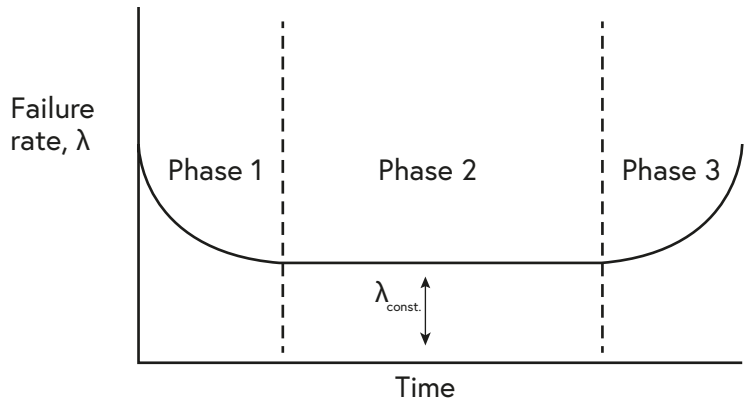


Figure 18.10 The bathtub curve is *wrongly* presented as universal

United Airlines discovered that the "bathtub curve" of failure rates only applied to about 4% of the aircraft components.[4] In particular, many exhibited no tendency to fail more when aging, which made policies of periodic replacement pointless. What the United people found was that aircraft components exhibited a variety of patterns and that 89% of the ones they analyzed never had a wear-out phase. As a consequence, there was no point in systematically replacing them after a fixed interval or use count.

Maintenance strategists in United Airlines needed a better model to maintain their aircraft. It was labeled "Reliability-Centered Maintenance" (RCM) in 1975 by the US Department of Defense. RCM is an ongoing, systematic process of matching critical systems with the most cost-effective maintenance strategy to maximize overall reliability. According to Anthony Smith, RCM rests on four steps:

4 Smith, A. M. (1993). *Reliability-Centered Maintenance*. McGraw-Hill.

1. Set the primary goal to preserve *system function*, not the function of every component.
2. Identify potential causes of *system failure*, focusing on the modes of the systems' components.
3. Since some failure modes are more critical than others, one should prioritize them, which is done using a decision tree logic.
4. Fourth, each failure mode is addressed by maintenance tasks that are effective – fixes the issue – and efficient – can be done at a reasonable cost. If the failure mode cannot be reasonably mitigated or is deemed to be of low priority, a "run-to-failure" strategy for the component may be used.

Since RCM allows a company to focus their maintenance strategy on the components that matters for system function, it is much more cost-efficient than replacing all parts at set intervals. RCM is still used in aircraft MRO (Maintenance, Repair, and Overhaul) (Figure 18.11). It has spread outside aircraft maintenance to other asset-intensive industries like nuclear power plants, railways, and process industries. Some RCM thinking has also found its way into machine-shops, for example in the form of redundant tools in machining center pockets. RCM is formalized in the standard SAE JA1011 by the Society of Automotive Engineers (SAE).

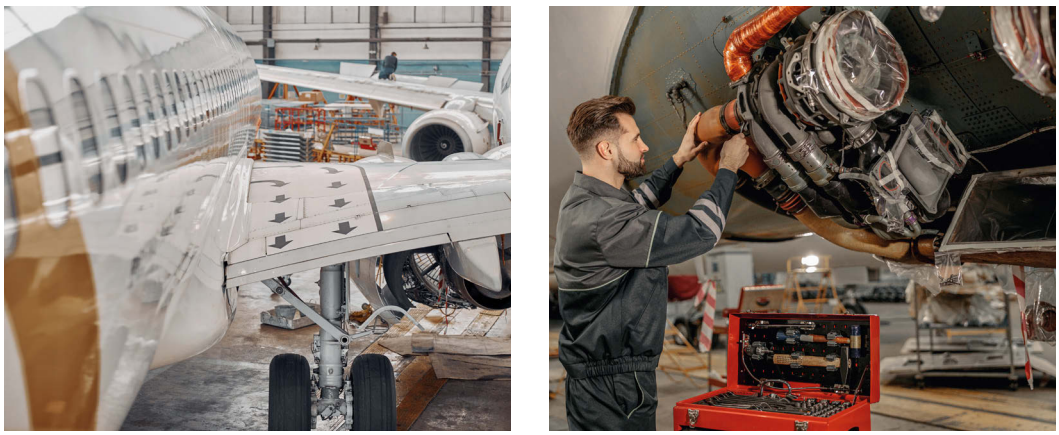


Figure 18.11 Aircraft maintenance

Stop and Think 18.4
Why is the aircraft industry a good place to look for the latest maintenance strategies?

Critics of RCM say that it is a workaround to the limitations of the equipment rather than an improvement of it. It is better to have a cutting tool that lasts twice as long than to put a redundant tool on standby in the machine, but then you have to find such a tool. Are failures not a problem when you have multiple, interchangeable machines with overcapacity? Logically, they are not, but it is another story from the human point of view. It won't be a problem next week, but what happens over time when overcapacity in an area allows you to have 25% of your equipment down? Your performance will eventually settle at a point where you actually have one machine in four down at any time. Why bother keeping all of them up all the time when they are not needed? Settling for this low availability, however, turns this process into a bottleneck.

Failure Mode and Effects Analysis

As part of RCM, United Airlines developed the technique Failure Mode and Effects Analysis (FMEA). The consequences of a component failure on an aircraft in flight also vary greatly depending on whether it is a passenger reading light, an avionic system, or the rudder. You don't need the reading light to stay in the air and you can't replace the rudder in flight, but you can have a standby avionic system take over. FMEA served as the basis for replacement policies, spare parts inventory management, and targeted equipment redundancies.

FMEA has become one of the most common methods of technical risk management. FMEA helps avoid critical dysfunction of products (machines) by diagnosing potential failure modes. "Failure modes" refers to the ways, or modes, in which something might fail. The"effects analysis" refers to studying the consequences of those failures.

The literature suggest many variants of methodological road maps for FMEA (7-step, 10-step, 15-step methods, etc.), but they all boil down to identifying and prioritizing potential failures according to how serious their consequences are, how frequently they occur, and how easily they can be detected before they occur. Based on this, failure modes are targeted for elimination or reduction according to their priority.

The FMEA concept is known in manufacturing, but more often applied to products in product development projects than to *production equipment*. FMEA has led to best practices such as targeted redundancies, for example, in machining centers by placing the same frequently used cutting tools in two pockets, with the second tool taking over when the first is worn out. The operator can then replace the worn tool while the machine is running with the fresh one.

18.3.4. **Predictive and condition-based maintenance**

Periodic preventive maintenance is either based on time – like yearly shutdowns on chemical reactors – or usage – like replacement of tools or dies are a given number of uses. *Predictive*[5] or *condition-based* maintenance is instead triggered by observations of the equipment indicating that it will soon fail. The goal is to avoid shutting it down whether it needs it or not while avoiding failures.

Periodic maintenance based on time or usage is supposed to be based on an understanding of the equipment acquired through training and experience that is baked into a schedule of actions; predictive or condition-based maintenance, on the other hand, is based on the real time status of the equipment. It requires monitoring vital signs with sensors, checking in real time the evolution of the measurements against a model of the equipment to detect trends and issue alarms. As of 2022, it is technically possible but not a common practice.

Many view *predictive* and *condition-based* maintenance as identical and use the names interchangeably. Other specialists, like Olga Fink,[6] make a sharp distinction:

5 Selcuk, S. (2017). Predictive maintenance, its implementation and latest trends. Proceedings of the Institution of Mechanical Engineers, Part B: *Journal of Engineering Manufacture, 231*(9), 1670–1679.

6 Fink, O. (2020). Data-driven intelligent predictive maintenance of industrial assets. In: Smith, A. (Ed.) *Women in Industrial and Systems Engineering*. Springer.

- **Condition-based maintenance** only considers the current state. It's fault detection and diagnosis based on the condition of the machine, as observed through routine inspection or functional tests on characteristics like oil quality. It's a threshold approach that doesn't require extensive data.
- **Predictive maintenance** prescribes actions based on a prediction of *remaining time to failure*, or remaining useful life. It is a more recent development than condition-based maintenance and sometimes called "early fault detection."

Condition-based maintenance is widely used in manufacturing; *predictive* maintenance, much less. If the prediction can be based on a linear model, it is straightforward; otherwise, it is a technical challenge including machine learning. Many manufacturers consider predictive maintenance to be too expensive, especially when applied to hundreds of machines.

Predictive maintenance monitors multiple observations on the equipment:

- **Vibrations**, on machines with rotating or reciprocating components
- **Oils and lubricants**, for consumption and composition while in use
- **Acoustic noise and ultrasound** from what people hear to what sensors detect
- **Heat** from thermometer readings at various locations on the machine to heat maps
- Other variables

In 2022, there are companies – including giant players like Siemens and Amazon and an array of start-ups – selling smart sensors, sensor kits, and machines equipped with sensors connected to the control system to support predictive maintenance. Data sharing across machines and, in particular, across companies, remains a show-stopper and limits large-scale adoption.

The next step after predictive maintenance is *prescriptive* maintenance, awkwardly abbreviated as "RxM." Prescriptive maintenance does not only predict future failures, but also suggests how to mitigate them. It uses machine learning on historical and near-real-time data to suggest or automatically adjust operating conditions that prolong the remaining useful life of acquirements. Prescriptive maintenance requires smart sensors, data storage, cloud solutions, analytics, and software. It is part of the "Industry 4.0" vision of self-optimizing factories.

18.3.5. Mean Time Between Failures and Mean Time To Repair

How long, on the average, can we run a machine before we need to repair it? How long does it take to repair it? These are obviously essential questions to ask when planning maintenance for repairable equipment. They are usually answered by the following indicators:

- **Mean Time Between Failures** (MTBF) is the mean time between consecutive failures of a piece of machinery during normal operations.
- **Mean Time To Repair** (MTTR) is the mean time from failure notification to return of the equipment to production status. The R in MTTR is sometimes also used for "Respond," "Recover," or "Resolve."

If we have a history of failures and repairs for a machine, we can simply take averages:

$$MTBF = \frac{Uptime}{Number\ of\ failures}$$

$$MTTR = \frac{Sum\ of\ repair\ times}{Number\ of\ repairs}$$

With these figures, we can calculate the uptime ratio:

$$Uptime\ ratio = \frac{MTBF}{MTBF + MTTR}$$

While these formulas are self-explanatory and helpful, they must be used with caution. They are based on historical data and averages, and therefore based on the assumption that equipment performance has been stationary through the historical period used. One characteristic of MTBFs and MTTRs that is not often discussed is that these metrics are in fact estimators of parameters in models of equipment reliability and queueing models of maintenance operations, as discussed below. Before drawing any conclusions from MTBF and MTTR data, one must know *exactly* how they were calculated and be wary of comparing values between factories where it may be different.

One caveat with the use of MTBFs with *production* equipment is that the volume of work it does usually is more relevant than the passage of time. An air conditioner on a roof will rust over time regardless of production volume inside the building but, in production, if a die is good for 6,000 shots, it doesn't matter whether the number is reached in a week or a month. If the volume of work is consistent or the failures are independent of it, you can use the MTBF.

> **Stop and Think 18.5**
> A chocolate packaging line runs normal operations for 100 hours before it breaks down. A motor is then replaced during 1 hour of maintenance. What is the MTBF and MTTR? After restarting the machine, it breaks down after 1 hour and it takes 3 hours to bring it back to operating condition. What is the MTBF and MTTR now? The company starts a preventive maintenance program that enables the line to run for 10,000 hours before the next failure and it takes 1 hour to repair it. What is the MTBF and MTTR now? What does this example tell us about MTBF and MTTR?

18.3.6. Probabilities of equipment system failure

A common model is that machines are systems of components that fail independently of one another and such that the failure of any one component is a failure of the whole system. The failure process of the machine is therefore the pooled output of the failure processes of its independent components. Regardless of the specific failure processes of the components, pooling them results in a constant failure rate λ for the machine as a whole, with the number of failures in disjoint intervals being independent random variables.

With a constant rate λ the probability of having one failure between t and $t+dt$ is λdt. If we split the interval *[0,t]* into n equal increments of length t/n , the probability of having no failure in *[0,t]* is

$$\lim_{n\to\infty}\left(1-\lambda\frac{t}{n}\right)^{n}=\exp(-\lambda t)$$

And the probability of having a time between failures of t is

$$h(t)=\exp(-\lambda t)\times\lambda$$

The expected value of the time to failures when working is therefore

$$E(MTBF)=\frac{1}{\lambda}$$

Calling it *time between failures* is a bit misleading, because it does not include the repair time. The time to repair is also often modeled as exponentially distributed with rate but the logic for doing so is not quite as compelling as for failures. Repairs range in scope from flicking a switch to replacing a major subsystem with a long delivery lead time. Sometimes technicians are immediately available to work on the repair; other times, there is a wait before they can even start… The notion that the machine comes back up between t and t+dt with probability μdt is convenient but simplistic.

In particular, it yields

$$E(MTTR)=\frac{1}{\mu}$$

In this model, if the machine is up, it fails with probability λdt between t and $t+dt$ regardless of how long it has been up or anything else in its past; if it's down, it comes up with probability μdt between t and $t+dt$ regardless of how long it's been down. In other words, given the present state, the future is independent of the past. It's called the Markov property, and it simplifies the math so much that it is tempting to ignore the less than perfect fit of repair times to this model.

It's simple for 1 machine but more interesting when you apply this thinking to a group of machines. From the perspective of one technician looking after n machines, it looks like a queueing system with the technician as the server and the machines that are down as the customers. One failure occurs among the n machines at the rate $\lambda=\lambda_1+\ldots+\lambda_n$, where λ_i is the failure rate of the i-th machine.

If the machines all have the same repair rate μ this is the classic M/M/1 queue discussed in the introduction of every text on queueing theory. The ratio $\rho=\lambda/\mu$ is the technician's utilization and $1-\rho$ is his or her availability. The mean number of machines that are down is

$$L=\frac{\rho}{1-\rho}$$

By Little's law, L = λW, and the MTTR for the shop, including waiting, is

$$MTTR=W=\frac{1}{\mu-\lambda}=\frac{1}{\mu\times(1-\rho)}$$

If, instead of one technician, you have a team of c technicians with equal skills looking after the n machines, you have an M/M/c queue, with more complex formulas for L and W that still grow to infinity like 1/(1-ρ) when the technicians' utilization ρ approaches 100%.

On the other hand, when ρ is low,

$$MTTR \approx \frac{1}{\mu}$$

as with one technician looking after one machine.

Perhaps the most interesting result from queueing theory in practical terms is Kingman's formula, according to which this saturation pattern occurs in queues called G/G/1, single-server queues where failure and repair times are only required to have means and standard deviations. In a broad array of circumstances, when technician utilization is high, the mean waiting time and queue length are roughly inversely proportional to technician availability. This is useful in assessing the value of increasing availability from, say, 1% to 5%.

In the discussion of the size of the Maintenance department, the objective was to have an empty queue, so that a technician is available whenever needed. This did not require concepts like MTBF or MTTR. They come into play when you are not in this situation.

18.3.7. Equipment maintenance cost versus replacement cost

Assume the plant has a fleet of 15 aging forklift trucks, of which only 5 are working at any time (Figure 18.12). According to Accounting, the yearly maintenance cost of this fleet is 80% of what it would take to buy 5 new forklift trucks that would require little or no maintenance for at least two years. Then there is a strong case for investing in a new fleet.

.



Figure 18.12 When to keep maintaining old forklifts versus buying new ones?

Generalizing from this example, we identify the ratio of yearly *maintenance* cost to *replacement* cost as a metric for each machine in the shop. Calculating it, however, is not as straightforward as the forklift example makes it sound:

- **Maintenance cost issues.** While undergoing maintenance, a machine unavailable for production. It consumes the time of production operations and maintenance technicians, uses equipment that is shared over time with other machines, and possibly involves sending the machine out to an outside shop. Consequently, the machine's maintenance cost, as calculated by Accounting, involves resource allocations among machines and over time, which accountants do in a variety of ways. In particular, this "maintenance cost" does not amount to cash that you save by replacing the machine. For each machine, this requires an analysis of the specific funds flows that would stop.
- **Replacement cost issues.** The replacement cost is not straightforward either. The first question is what to replace the old machine with. Replacing it with another machine of the same vintage is unlikely to reduce maintenance requirements. Buying a brand new machine, on the other hand, is likely to be an upgrade rather than just a replacement, which raises the question of the value to the plant of its additional capabilities. In some cases, like IT, lightly used equipment can be a better choice than new equipment. The second issue is that there is more to replacing equipment than purchasing it. You have to install it, train operators, technicians, and engineers to use it, and start it up.

The end result is that a retain-or-replace decision must be based on more than an analysis of a single metric. In addition to the costs, one must also consider user experience, safety issues, supplier availability, and compatibility issues, among other things.

Summary

The most essential points covered in this chapter were as follows:

1. Maintenance is essential to a well-functioning factory.
2. The two overall maintenance strategies are preventive versus corrective maintenance.
   - Preventive maintenance covers periodic, condition-based, and predictive maintenance.
   - Corrective maintenance covers run-to-failure and unplanned failure.
3. CMMS help keep track of workorders, spare parts inventory, and other maintenance-related tasks.
4. 5S helps maintain an uncluttered, clean, and tidy environment for production equipment, which eliminates many potential causes of equipment problems.
5. Autonomous Maintenance delegates routine tasks like changing light bulbs and frayed cables to production operators.
6. Formulas and queuing theory help plan Mean Time Between Failures (MTBF) and Mean Time To Repair (MTTR), but must be interpreted carefully since the calculations are based on past data and averages.
7. Total Productive Maintenance (TPM) has evolved into a full-scale production improvement program for process industries.
8. Reliability-Centered Maintenance (RCM) reduces maintenance costs by focusing attention on system-critical components.
9. Predictive maintenance is starting to find applications in industry, while prescriptive maintenance is on the horizon.
10. Replace machines whose yearly maintenance costs exceeds their replacement cost.

Test yourself

1. What is the contribution of maintenance to manufacturing?
2. What are the main functions of a computerized maintenance management system (CMMS)?
3. How does condition-based maintenance differ from periodic maintenance?
4. What is predictive maintenance?
5. What is prescriptive maintenance
6. What is run-to-failure maintenance?
7. What is 5S and what does it contribute to maintenance?
8. What is the purpose of an FMEA analysis?

9 What is the benefit of Autonomous Maintenance?

10 What is Mean Time Between Failures?

Further reading

Bokrantz, J., Skoogh, A., Berlin, C., Wuest, T., & Stahre, J. (2020). Smart maintenance: A research agenda for industrial maintenance management. *International Journal of Production Economics, 224*, 107547.

Garg, A. & Deshmukh, S. G. (2006). Maintenance management: Literature review and directions. *Journal of Quality in Maintenance Engineering, 12*(3), 205–238.

Levitt, J. (2005). Managing factory maintenance. Industrial Press Inc.

Mobley, R. K. (2002). An introduction to predictive maintenance (2nd ed.). Butterworth-Heinemann.

Moore, R. (2004). *Making common sense common practice*. Elsevier.

Nachi-Fujikoshi Corporation (1990). *Training for TPM*. Productivity Press.

Nakajima, S. (1988). *Introduction to TPM*. Productivity Press.

Smith, A. M. (1993). *Reliability-Centered Maintenance*. McGraw-Hill.

Suehiro, K. (1992). *Eliminating minor stoppages on automated lines*. Productivity press.

Williams, J. H., Davies, J. H., & Drake, P. R. (Eds.) (1994). *Condition-based maintenance and machine diagnostics*. Springer.



Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Bibliography

Abrahamson, E. (1991). Managerial fads and fashions: The diffusion and rejection of innovations. *Academy of Management Review, 16*(3), 586–612.

Acton, P. H. & Acton, P. (2014). *Poiesis: Manufacturing in Classical Athens*. Oxford University Press.

Agricola, G. (2013). *De Re Metallica*. Dover Publications.

Aguilar, F. J. (1967). *Scanning the business environment*. Macmillan.

Anderson, K. (2020). *Learning to lead, leading to learn: Lessons from Toyota leader Isao Yoshino on a lifetime of continuous learning*. Integrand Press.

Antony, J., McDermott, O., & Sony, M. (2021). *Revisiting Ishikawa's original seven basic tools of quality control: A global study and some new insights*. IEEE Transactions on Engineering Management. In press.

Aoudia, H. (2015). *Perfect QRQC: Volume 2, Prevention, standardization, coaching*. Maxima.

Aoudia, H. & Testa, Q. (2012). *Perfect QRQC (Quick Response Quality Control)*. Maxima.

Arai, K. & Sekine, K. (2006). *Kaizen for quick changeover: Going beyond SMED*. Taylor & Francis.

Astakhov, V. P. (2012). Design of experiment methods in manufacturing: Basics and practical applications. In *Statistical and computational techniques in manufacturing* (pp. 1–54). Springer.

Babst, A. & Sander, M. (2021, March 12). 1320 iPhones pro Schicht (p. 23.). *Neue Zürcher Zeitung*.

Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120.

Beck, K., Beedle, M., van Bennekum, A., et al. (2001). *The Agile Manifesto*. Agile Alliance. URL: http://agilemanifesto.org.

Bellgran, M., Höckerdal, K., Kruvde, M., & Wiktorsson, M. (2012). *Green Performance Map handbook*. Mälardalens Högskola.

Benner, M. J. & M. L. Tushman. (2003). Exploitation, exploration, and process management: The productivity dilemma revisited. *Academy of Management Review, 28*(2), 238–256.

Bicheno, J. & Holweg, M. (2009). *The lean toolbox*. PICSIE Books.

Biringuccio, V. (1990). *The pirotechnia of Vannoccio Biringuccio*. Dover Publications.

Blauner, R. (1964). *Alienation and freedom: The factory worker and his industry*. University of Chicago Press.

Brandenburger, A. & Nalebuff, B. (1996). *Co-opetition*. Currency/Doubleday.

Brooks, F. (1975). *The mythical man-month*. Addison–Wesley.

Browning, T. R. & Eppinger, S. D. (2016). *Design structure matrix methods and applications*. MIT Press.

Burbidge, J. L. (1989). *Production flow analysis*. Oxford University Press.

Chandler Jr., A. D. (1977). *The visible hand: The managerial revolution in American business*. Harvard University Press.

Chesbrough, H. W. (2006). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.

China Daily (2017, September 19). Half of iPhones manufactured in central China's Zhengzhou city.

Christopher, M. (1998). *Logistics and supply change management* (2nd ed.). Pitman.

Clymer, B. (2015, March 16). *Inside the manufacture: Going where few have gone before – Inside all four Rolex manufacturing facilities*. HODINKEE. www.hodinkee.com.

Coase, R. H. (1937). The nature of the firm. *Economica, 4*(16), 386–405.

Cole, R. E. (2018). *Work, mobility, and participation*. University of California Press.

Crosby, P. (1979). *Quality is free*. McGraw–Hill.

Cusumano, M. A., Holweg, M., Howell, J., Netland, T., Shah, R., Shook, J., Ward, P., & Womack, J. (2021). Commentaries on "the lenses of lean". *Journal of Operations Management, 67*(5), 627–639.

Davis, M. D. (2012). *Game theory: A nontechnical introduction*. Dover Publications.

Demetrescoux, R. (2019). *La boîte à outils du lean* (2nd ed.). Dunod.

Deming, W. E. (1986). *Out of the crisis: Quality, productivity and competitive position*. Cambridge University Press.

Deming, W. E. (2000). *The new economics: For industry, government, education*. MIT Press.

Dennis, P. (2006). *Getting the right things done: A leader's guide to planning and execution*. Lean Enterprise Institute.

Diderot, D. (1993). *A Diderot pictorial encyclopedia of trades and industry*. Dover Publications.

Diebold, J. (1952). *Automation. The advent of the automatic factory*. D. Van Nostrand Company.

Doerr, J. (2017). *Measure what matters: How Bono, the Gates Foundation, and Google rock the world with OKRs*. Penguin Publishing Group.

Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review, 70*(11), 35–36.

Drucker, P. (1954). *The practice of management*. Harper/Heinemann. (2007). (Revised ed.). Butterworth–Heinemann.

Drucker, P. (1993). *Concept of the corporation*. Transaction Publishers.

Drucker, P. (2012). *Management*. Taylor & Francis.

Feigenbaum, A. V. (2008). *Total quality control*. I.K. International Publishing House.

Ferdows, K. (1997). Made in the world: The global spread of production. *Production and Operations Management, 6*(2), 102–109.

Ferdows, K. & De Meyer, A. (1990). Lasting improvements in manufacturing performance: In search of a new theory. *Journal of Operations Management, 9*(2), 168–84.

Fine, C. (1998). *Clockspeed: Winning industry control in the age of temporary advantage*. Perseus.

Fink, O. (2020). Data-driven intelligent predictive maintenance of industrial assets. In Smith, A. (Ed.) *Women in Industrial and Systems Engineering*. Springer.

Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd.

Fiume, O. J. & Cunningham, J. E. (2017). *Real numbers: Management accounting in a lean organization*. Jean Cunningham Consulting.

Forrester, J. W. (1961). *Industrial dynamics*. MIT Press.

Frandson, A. (2019). *Takt time planning as a work structuring method to improve construction work flow* [PhD Dissertation, Civil and Environmental Engineering, UC Berkeley]. URL: https://escholarship.org/uc/item/6dp4n4fz.

Franklin, B., Franklin, W. T., & Duane, W. (1834). *Memoirs of Benjamin Franklin*. M'Carty & Davis.

Fretty, P. (2020, December 10). Stanley Black & Decker realizes AI is a powerful tool. *Industry Week*.

Friedli, T. & Bastoen, B. (2006). *Operational excellence in the pharmaceutical industry*. ECV, Editio Cantor Verlag.

Fujimoto, T. (1999). *The evolution of a manufacturing system at Toyota*. Oxford University Press.

Gilbreth, F. & Gilbreth L. (1921). *Process charts.* Annual Meeting of the American Society of Mechanical Engineers, December 5–9, New York.

Goldratt, E. M. (1997). *Critical chain*. North River Press.

Grant, R. M. (1991). The resource-based theory of competitive advantage: Implications for strategy formulation. *California Management Review, 33*(3), 114–135.

Grove, A. S. (1995). *High output management*. Random House. (Original work published in 1983).

Grove, A. S. (2010). *Only the paranoid survive: How to exploit the crisis points that challenge every company*. Crown.

Gryna, F. M. & Juran, J. M. (1988). *Juran's quality control handbook*. McGraw–Hill.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR), 51*(5), 1–42.

Hall, R. W. (1983). *Zero inventories*. Dow Jones-Irwin.

Hamel, G. & Prahalad, C. K. (1990). Core competence of the organization. *Harvard Business Review, 68*(3), 79–91.

Hammer, M. & Champy, J. (1993). *Business process reengineering*. Harper Collins.

Harry, M. & Lawson, R. (1990). *Six sigma producibility analysis and process characterization* (Publication Nr. 60-3-03/88). Motorola University Press.

Harry, M. J. & Schroeder, R. (2000). *Six sigma: The breakthrough management strategy revolutionizing the world's top corporations*. Currency.

Hayes, R. H. & Wheelwright, S. C. (1979). Link manufacturing process and product life cycle. *Harvard Business Review, 57*(1), 133–140.

Hayes, R. H. & Wheelwright, S. C. (1984). *Restoring our competitive edge: Competing through manufacturing*. John Wiley & Sons, New York.

Heinrich, H. W. (1931). *Industrial accident prevention: A scientific approach*. McGraw–Hill.

Hillier, F. S. & Lieberman, G. J. (2015). *Introduction to Operations Research*. McGraw–Hill.

Hinckley, C. M. (2001). *Make no mistake! An outcome-based approach to mistake-proofing*. Taylor & Francis.

Hines, P., Found, P., Griffiths, G., & Harrison, R. (2008). *Staying lean: Thriving not just surviving*. Lean Enterprise Center, Cardiff University.

Hounshell, D. (1995). Ford automates: Technology and organization in theory and practice. *Business and Economic History, 24*(1), 69.

Imai, M. (1986). *Kaizen (Ky'zen), the key to Japan's competitive success*. Random House.

Ishikawa, K. (1968). *Guide to quality control*. Asian Productivity Organization, Tokyo.

Ishikawa, K. (1985). *What is total quality control? The Japanese way*. Prentice-Hall.

John, D. (1952). *Automation. The advent of the automatic factory*. D. Van Nostrand Company.

Juran, J. M. & Gryna, F. M. (Eds.). (1988). *Juran's quality handbook: The complete guide to performance excellence* (4th ed.). McGraw–Hill Education.

Kadoma, H. (2019). *Poka-yoke countermeasures: Preventing human error! Human error countermeasures*. Monozukuri Library.

Kaplan, R. S. & Norton, D. P. (1996). Strategic learning & the balanced scorecard. *Strategy & Leadership, 24*(5), 18–24.

Kapp, K. M. (1997). The USA Principle. *APICS –The Performance Advantage, 7*, 62–66.

Kelley, J. E., Walker, M. R., & Sayer, J. S. (1989). The origins of CPM: A personal history. *PM Network, 3*(2), 7–22.

King, G. (2008, April 13). Tesla CEO Elon Musk, stressed but "optimistic," predicts big increase in Model 3 production. *CBS Morning News*. URL: https://www.cbsnews.com/news/elon-musk-tesla-model-3-problems-interview-today-2018-04-13/.

Kingman, J. F. C. (1961). The single server queue in heavy traffic. *Mathematical Proceedings of the Cambridge Philosophical Society, 57*(4), 902.

Kipling, R. (2009). *Just so stories for little children*. Oxford University Press. (Original work published 1902.)

Kolesar, P. (1994). What Deming told the Japanese in 1950. *Quality Management Journal, 2*(1), 9–24.

Krafcik, J. F. (1988). Triumph of the lean production system. *Sloan Management Review, 30*(1), 41–51.

Krajewski, M. (2014). The great lightbulb conspiracy. *IEEE Spectrum, 51*(10).

Kurvde, M. (2014). *Development of collaborative green lean production systems* (No. 162). [Doctoral Thesis, Mälardalen University]. Mälardalen University Press Dissertations.

Lambert, D. M. & Cooper, M. C. (2000). Issues in supply chain management. *Industrial Marketing Management, 29*(1), 65–83.

Laraia, A. C., Moody, P. E., & Hall, R. W. (1999). *The kaizen blitz: Accelerating breakthroughs in productivity and performance*. Wiley.

Lee, H. (1997). The bullwhip effect in supply chains. *MIT Sloan Management Review, 38*(3), 93–102.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management Science, 43*(4), 546–558.

Lee, J. (2020). *Industrial AI: Applications with sustainable performance.* Springer.

Lepore, J. (2019, March 4). The robot caravan. *New Yorker*. URL: https://www.newyorker.com/magazine/2019/03/04/.

LightGBM's documentation. URL: https://lightgbm.readthedocs.io.

Liker, J. K. & Hoseus, M. (2008). *Toyota culture*. McGraw–Hill Publishing.

Ling, R. C. (1988). *Orchestrating success: Improve control of the business with sales & operations planning*. J. Wiley.

Little, J. D. C. (1961). A proof for the Queuing Formula: $L = \lambda W$. *Operations Research, 9*(3), 383–387.

Magee, D. (2003). *Turnaround: How Carlos Ghosn rescued Nissan*. Harper Business.

Maghazei, O., Lewis, M., & Netland, T. (2022). Emerging technologies and the use case: A multi-year study. *Journal of Operations Management, 68*(6–7), 560–591.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusions and way forward. *International Journal of Forecasting, 34*(4), 802–808.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2021). The M5 competition: Background, organization, and implementation. *International Journal of Forecasting*. In press.

Mann, D. (2014). *Creating a lean culture: Tools to sustain lean conversions*. CRC Press.

Margretta, J. (1998). Fast, global, and entrepreneurial: Supply chain management, Hong Kong style. *Harvard Business Review, 76*(5), 102–115.

May, G. S. & Spanos, C. J. (2006). *Fundamentals of semiconductor manufacturing and process control*. John Wiley & Sons.

Maynard, H. B., Stegemerten, G. J., & Schwab, J. L. (1948). *Methods-time measurement*. McGraw–Hill.

McDonald, C. J. (1998). The evolution of Intel's Copy EXACTLY! technology transfer method. *Intel Technology Journal, Q4*, pp. 1–6.

Mercier, J.-P. (1987). *La Maintenance des Centrales Nucléaires*. Kirk I.

Monden, Y. (2010). *Toyota Production System: An integrated approach to just-in-time*. Productivity Press.

Morgan, J. & Liker, J. K. (2020). *The Toyota Product Development System: Integrating people, process, and technology*. Taylor & Francis.

Muther, R. (1973). *Systematic layout planning*. Cahners Books.

Nagel, R. N. & Dove, R. (1991). *21st century manufacturing enterprise strategy: An industry-led view*. Diane Publishing.

Nakajima, S. (1988). *Introduction to TPM: Total Productive Maintenance*. Productivity Press.

Narusawa, T, & Shook, J. (2009). *Kaizen express: Fundamentals for your lean journey*. Lean Enterprise Institute

NBC (1980). If Japan can, why can't we? URL: https://www.youtube.com/watch?v=vcG_Pmt_Ny4.

Neave, H. R. (1990). *The Deming dimension*. SPC Press.

Netland, T. (2013). Exploring the phenomenon of company-specific production systems: One-best-way or own-best-way? *International Journal of Production Research, 51*(4), 1084–1097.

Netland, T. & Aspelund, A. (2014). Multi-plant improvement programmes: A literature review and research agenda. *International Journal of Operations & Production Management, 34*(3), 390–418.

Netland, T., Schloetzer, J. D., & Ferdows, K. (2015). Implementing corporate lean programs: The effect of management control practices. *Journal of Operations Management, 36*(0), 90–102.

Nikkan Kogyo Shimbun (Ed.). (1988). *Poka-yoke: Improving product quality by preventing defects*. Taylor & Francis.

Nishiguchi, T. & Beaudet, A. (1998). The Toyota group and the Aisin fire. *Sloan Management Review, 40*(1), 49–60.

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic Books.

OECD (2002). *Benchmarking intermodal freight transport*. Paris, Report number: E112021.

Ohno, T. (1988). *Toyota Production System: Beyond large-scale production*. Productivity Press.

Olhager, J. (2003). Strategic positioning of the order penetration point. *International Journal of Production Economics, 85*(3), 319–329.

Parker, D., Riley, K., Robinson, S., Symington, H., Tewson,J., Jansson, K., Ramkumar, S., & Peck D. (2015). *Remanufacturing market study*. European Remanufacturing Network. EU report, Reference: EC--09 404 ERN.

Pay, R. (2008, March 1). Everybody's jumping on the lean bandwagon, but many are being taken for a ride. *Industry Week*. URL: https://www.industryweek.com/leadership/companies-executives/article/21946187/everybodys-jumping-on-the-lean-bandwagon-but-many-are-being-taken-for-a-ride.

Penrose, E. T. (1959). *The theory of the growth of the firm*. Basil Blackwell.

PMI (2021). *Guide to the Project Management Body of Knowledge (PMBOK Guide) and the standard for project management*. Project Management Institute.

Porter, M. E. (1979). How competitive forces shape strategy. *Harvard Business Review 57*(2), 137–145.

Powell, D. J. & Coughlan, P. (2020). Rethinking lean supplier development as a learning system. *International Journal of Operations & Production Management*, 40(7/8), 921–943.

Ptak, C. A. & Smith, C. (2019). *Demand Driven Material Requirements Planning (DDMRP): Version 3.* Industrial Press, Inc.

Rauwendaal, C. (2014). *Polymer extrusion*. Hanser/Gardner Publications.

Rauwendaal, C. (2019). *Understanding extrusion*. Hanser/Gardner Publications.

Repenning, N. P. & Sterman, J. D. (2001). Nobody ever gets credit for fixing problems that never happened: Creating and sustaining process improvement. *California Management Review, 43*(4), 64–88.

Rosenberg, J. & Schmidt, E. (2017). *How Google works*. Grand Central Publishing.

Rother, M. & Shook, J. (2003). *Learning to see: Value-stream mapping to create value and eliminate Muda*. Lean Enterprise Institute.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*, 206–215.

Rülein von Calw, U. (1955). *Ulrich Rülein von Calw und sein Bergbüchlein* (Pieper, W., Ed.). Akademie-Verlag. (Original work published in 1505.)

Salvendry, G. (Ed.). (2007). *Handbook of Industrial Engineering*. John Wiley.

Santos-Fernández, E. & Scagliarini, M. (2012). MPCI: An R package for computing multivariate process capability indices. *Journal of Statistical Software, 47*(7), 1–15.

Scheer, A.-W. (1998). *Business process engineering study edition: Reference models for industrial enterprises* (pp. 120–121). Springer Verlag.

Schonberger, R. (1982). *Japanese manufacturing techniques: Nine hidden lessons in simplicity*. The Free Press.

Schumpeter, J. A. (1942). *Capitalism, socialism, and democracy*. Harper & Brothers.

Schwab, K. (2016). *The fourth industrial revolution*. World Economic Forum, Geneva, Switzerland.

Selcuk, S. (2017). Predictive maintenance, its implementation and latest trends. Proceedings of the Institution of Mechanical Engineers, Part B. *Journal of Engineering Manufacture, 231*(9), 1670–1679.

Senoner, J., Netland, T., & Feuerriegel, S. (2022). Using explainable artificial intelligence to improve process quality: Evidence from semiconductor manufacturing. *Management Science, 68(*8), 5704–5723.

Shah, R. & Ward, P. T. (2003). Lean manufacturing: Context, practice bundles, and performance. *Journal of Operations Management, 21*(2), 129–149.

Shah, R. & Ward, P. T. (2007). Defining and developing measures of lean production. *Journal of Operations Management, 25*(4), 785–805.

Shewhart, W. A. (1939). *Statistical method from the viewpoint of quality control*. Dover Publications.

Shingo, S. (1985). *A revolution in manufacturing: The SMED system*. CRC Press.

Shingo, S. (1986). *Zero quality control: Source inspection and the poka-yoke system*. Productivity Press.

Shingo, S. & Dillon, A.P. (1989). *A study of the Toyota Production System: From an Industrial Engineering viewpoint*. Productivity Press.

Shook, J. (2008). *Managing to learn: Using the A3 Management Process to solve problems, gain agreement, mentor and lead*. Lean Enterprise Institute.

Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E., & Shankar, R. (2008). *Designing and managing the supply chain: Concepts, strategies and case studies*. Tata McGraw–Hill Education. Chapter 4. The Barilla SpA case study.

Simchi-Levi, D., Schmidt, W., Wei, Y., Zhang, P. Y., Combs, K., Ge, Y., Gusekhin, O., Sanders, M., & Zhang, D. (2015). Identifying risks and mitigating disruptions in the automotive supply chain. *Interfaces, 45*(5), 375–390.

Skinner, W. (1969). Manufacturing: The missing link in corporate strategy. *Harvard Business Review, 47*(3), 79–91.

Skinner, W. (1974). The focused factory. *Harvard Business Review, 52*(3), 113–121.

Skinner, W. (1985). *Manufacturing: The formidable competitive weapon.* Wiley.

Sloan, A. (1990). *My years with General Motors*. Currency.

Smith, A. (1904). *The Wealth of Nations* (Cannan, E., Ed.). Methuen & Co. (Original work published in 1776).

Smith, A. M. (1993). *Reliability-centered maintenance*. McGraw–Hill.

Smyl, S., Rangagathan, J., & Pasqua, A. (2018). *M4 forecasting competition: Introducing a new hybrid ES-RNN Model.* Uber Engineering.

Spearman, M., Woodruff, D., & Hopp, W. (1990). CONWIP: A pull alternative to kanban. *International Journal of Production Research, 28*(5), 879–894.

Stalk, G. & Hout, T. (1990). *Competing against time: How time-based strategies deliver superior performance.* The Free Press.

Suri, R. (1998). *Quick response manufacturing: A companywide approach to reducing lead times*. Productivity Press.

Suzaki, K. (1987). *New manufacturing challenge: Techniques for Continuous Improvement*. The Free Press.

Taylor, F. W. (1911). *The principles of Scientific Management*. Harper & Brothers.

Toffler, A. (1970). *Future shock*. Random House Publishing Group.

Toyoda, E. (1983). *Creativity, challenge, and courage*. Toyota Motor Corporation, Internal document.

Toyota Motor Corporation (2001). *The Toyota Way 2001*. Internal document.

Tozawa, B. & Bodek, N. (2001). *The idea generator: Quick and easy kaizen*. PCS Press.

Tufte, E. (1990). *Envisioning information.* Graphics Press.

Tufte, E. (2003). *The cognitive style of Powerpoint*. URL: https://www.edwardtufte.com/tufte/powerpoint.

Turing, A. (1947, Feburary 20). *Lecture to the London Mathematical Society*. URL: https://www.vordenker.de/downloads/turing-vorlesung.pdf.

Tushman, M. L. & O'Reilly, C. A. (1996). The ambidextrous organization: Managing evolutionary and revolutionary change. *California Management Review, 38*(4), 1–23.

U.S. Bureau of Labor Statistics (2018). *All employees: Manufacturing* [MANEMP]. retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/MANEMP

Ulrich, K. T. & Eppinger, S. D. (1995). Design for manufacturing. In *Product Design and Development*. McGraw–Hill.

United Nations (1990). *Standard industrial classification of all economic activities*. Revision 3, Series M, No. 4, Rev. 4. URL: https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf.

van der Aalst, W. (2016). *Process mining: Data science in action*. Springer.

Vázsonyi, A. (1954). The use of mathematics in production and inventory control. *Management Science, 1*(1), 70–85.

VDI (2007). *Fertigungsmanagementsysteme: Manufacturing Execution Systems (MES)*. VDI 5600.

Warren, M. (2010). *The TWI Report*. Lulu Press, Inc.

Westerman, G., Bonnet, D., & McAfee, A. (2014). The nine elements of digital transformation. *MIT Sloan Management Review, 55*(3), 1–6.

Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. Macmillan.

Womack, J. P. & Jones, D. T. (1996). *Lean thinking: Banish waste and create wealth in your corporation.* The Free Press.

Womack, J. P., Jones, D. T., & Roos, D. (1990). *The machine that changed the world.* Rawson Associates.

World Bank (2018). *World Bank national accounts data, and OECD National Accounts data files*. URL: https://data.worldbank.org/.

Zandin, K. B. (2003). *MOST work measurement systems*. Marcel Dekker.

Zeccardi, J. (1988). Inspection and test. In Juran, J. M. & Gryna, F. M. (Eds.). *Juran's quality handbook: The complete guide to performance excellence* (4th ed., Chapter 18). McGraw–Hill Education.

Zehtaban, L. & Roller, D. (2015). Automated rule-based system for Opitz feature recognition and code generation from STEP. *Computer-Aided Design and Applications, 13*(3), 1–11.

Zinkgraf, S. A. (2006). *Six sigma – The first 90 days*. Pearson Education.

Index

3D printing *see* additive manufacturing
3PL *see* third-party logistics
4A framework 680
4PL *see* fourth-party logistics
5G 290, 293
5S 678, 740, 746, 749–753
8D model 602, 724

A

A3 599, 621, 655
ABC *see* Activity-Based Costing
ABC categorization 231, 275, 435–436
ABC/XYZ 275
acceptance sampling 688–689, 701–702
Activity-Based Costing 613–614
additive manufacturing 126, 147–150, 290, 495–497, 633
Advanced Planning and Scheduling 34, 102, 290, 350
advance shipping notice 370, 573
affordance 225
after-sales service 58–59, 148
AGCO 34, 402
agility 55, 58, 691
AGV *see* Automatic Guided Vehicle
Aquilar, F. 50
AI *see* artificial intelligence
Airbus 79, 658
Aisin Seiki fire 379, 627, 630–631
Akaike Information Criterion 84
Alibaba 87, 584
alienation 43, 218–219
all-hands meeting 548
Allied Signal 711
Amazon 87, 212, 223, 298, 299, 305, 317, 384, 493, 513, 517, 540, 584, 585, 676, 757
AME *see* Association for Manufacturing Excellence
American National Standards Institute 262, 664
American System of Manufacture 21, 27, 29, 41
analysis of variance 135, 139, 707
andon 311, 364–366
andon board 365
andon light 273, 311 *see also* tower lights
andon rope 365
ANOVA *see* analysis of variance
ANSI *see* American National Standards Institute
API *see* Application Programming Interface
APICS 559
Apple 59, 180, 581–582, 584–585
Application Programming Interface 293, 305
Applied Materials 482, 575
apprenticeship 23
APS *see* Advanced Production Scheduling
ARIMA *see* Autoregressive Moving Average
Arkwright, R. 24
ARMA *see* Autoregressive Moving Average
ARRIVAL 270
artificial intelligence 219, 223, 290–291, 530, 648, 691, 712–715
AS/RS *see* Automated Storage and Retrieval System
Asimov, I. 211
ASME symbols 117–118, 277
ASQ 725
assemble-to-order 65, 68
assembly 155–194
assembly cell 170–175, 241, 464–465, 483
assembly instructions 390–391 *see also* operator instructions; operator job design; work combination chart
assembly line 28–30, 106, 125–126, 156, 159–169, 192, 208–209, 241–243, 344, 432–433, 731–733
assembly station 176–186
Association for Manufacturing Excellence 642
as-you-go switch 208–210
Audi 169, 173
auto-ID *see* automatic identification
automated assembly 186, 311, 718
Automated Storage and Retrieval System 268, 536–540
Automatic Guided Vehicle 261, 323, 423, 463–464, 555
automatic identification 290, 307, 310, 314, 490, 548 *see also* barcode; datamatrix; QR code; radio frequency identification
automatic loading 210, 213–214, 228
automatic loom 196, 199
automatic palletizing 426, 459–460
automatic stop 52, 213, 216
automatic transfer 215, 305, 326
automatic unloading 208–210, 213–214 *see also* chaku-chaku
automation 32, 35, 186, 195–234, 458–460, 536–540
autonomation 199
autonomous maintenance 750, 753, 762
autonomy 172
autoregression 82, 84
autoregressive moving average 82–84
Azure 298

B

balanced scorecard 596
barcode 222, 290, 307, 314 490–493, 533
Barilla 74, 568
batch processes 126, 129–130
batch size 94–95, 271, 339–341, 608
bathtub curve 754
baton-touch 174, 193
the beer game 566
bench assembly 161–163, 243
benchmarking 49, 600, 604, 623, 667
Bethlehem Steel Works 30
Bicheno, J. 650
big data 299–300, 308
bill of materials 114, 133, 157, 176, 368–371, 378, 390–409
Black Belts 711
Blanc, H. 27
blanket purchase order 334–335, 360
Blauner, R. 219
block stacking 414, 527, 536, 538
BMW 186, 203
Bodek, N. 641, 667
body-in-white assembly 211, 270, 277, 733
body-on-sequence 336, 347, 506, 516, 549–550, 570–571
Boeing 737 54, 79–80, 99–100, 248, 733
BOM *see* bill of materials
bottleneck 77, 142, 147, 268, 539–540, 755
BPR *see* Business Process Re-engineering
brainstorming 611, 649, 655–656, 728
Brandenburger, A. 53-54, 334, 577–578
brand management 48

Bridgestone 599
Brooks, F. 655
Brother Industries 631
brownfield 255
bucket brigade 110, 174–175
bulk item 156–157, 305
bulletin board 595, 621
bullwhip effect 74, 336, 557, 564–566, 569, 571
Burbidge, J. L. 281
business intelligence 85, 316
Business Process Re-engineering 666
business strategy 43–45, 47, 730

C

Caberlon, C. 402
CAD *see* computer-aided design
capability index 135–137, 140
capacity planning 90–103
Čapek, K. 205
car assembly 104, 160, 167–168, 278, 344, 516, 549–550, 560, 571, 743
card-based pull methods 353–363
cardboard engineering 121, 177, 282–283, 286
cardboard mockups *see* cardboard engineering
Çatalhöyük 21
cause and effect diagram 650 *see also* fishbone diagram
CBM *see* Condition-Based Maintenance
cell assembly *see* assembly cell
cell *see* assembly cell; machining cell
cell layout 140–142, 170–175, 282
cellular manufacturing 174–175 *see also* assembly cell
chaku-chaku 206, 208–210, 213–214 *see also* automatic unloading
Chandler, A. 438
changeover 52, 341–342, 670 *see also* Single-Minute Exchange of Dies
Chaplin, C. 29, 163
check sheet 650, 729
China 22, 38, 40, 55, 155, 180, 202, 584
Choi, S. 60
Cicero 688
CIM *see* Computer-Integrated Manufacturing
cleanliness 18, 246, 477, 749–752
cleanroom 56, 261, 714
client/server model 293
clockspeed 54– 55, 330–331
closed-loop MRP 35, 368–369
closeness rating analysis 274, 278–280
CMMS *see* Computerized Maintenance Management Systems
CNC *see* Computer Numerical Controls
cobots 206, 212, 676 *see also* collaborative robots
Coca-Cola 60–61, 77, 81
CODP *see* customer order decoupling point
collaboration software 295, 655
collaborative supplier-customer relationship 571–580
collaborative robot 186, 206 *see also* cobots
common kanban 355–359
company-specific production system 34, 666, 673–676, 679–680
competitive advantage 45–47, 53, 57, 665
competitive capabilities 43, 47–49, 57, 690
competitive priorities 43, 47–48
complementors 53–54, 333–334, 378, 698
Computer-Aided Design 147–148, 258, 290, 309, 324, 371
Computer-Integrated Manufacturing 35, 676
Computerized Maintenance Management Systems 290–291, 411, 747–749
Computer Numerical Controls 114, 145–146, 213, 218–219, 290, 309–312
concurrent engineering 115, 144
Condition-Based Maintenance 739, 756–757
Cone, G. 711
consignment stocks 435, 550, 567
consolidation center 504–510, 552, 626
container 471–474, 476–490
continuous flow 126, 241, 458
continuous improvement 639–643, 652–653 *see also* Kaizen
contract manufacturer 51, 58–60, 67, 99, 149, 331
control chart 649–650, 689, 694, 705–706, 717
control panel 222, 225–226, 311
control room 226, 273
control theory 8, 140, 199–200
conveyor 179, 181, 289, 423, 456–461, 542
CONWIP 362–363
co-opetition 53–54, 578
Copy EXACTLY! 116, 416
core competency 61–62, 149
corporate strategy 44
corrective action 724, 727
cost centers 5, 67
COVID-19 38, 79–80, 255, 472, 624, 629–630, 633–634
C_p *see* capability index
C_{pk} *see* capability index
CRA *see* closeness rating analysis
craft production 21–23, 25, 41
crane 176, 207, 220, 229, 311, 460
Critical Path Method 658–660
CRM *see* customer relationship management
Crosby, P. 686
Crowther, S. 28
Culp, L. 711
customer order decoupling point 43, 65–68, 100
customer relationship management 290, 316, 580
customer-supplier relationship 558, 573–577, 627
cyber attack 300–301
cyber-physical systems 140, 292, 298, 304, 442–443
cybersecurity 300, 308, 677
cycle counting 335, 369, 532, 545–546, 608
cycle time 105, 130, 180, 604–605, 623

D

Dahl, R. 3
daily management meeting 617–619
data acquisition 307, 315
Database Management Systems 297
data lake 298
DataMatrix 308, 490–491
data mining 381, 385, 691
data science 280, 688, 690–692, 695, 705, 712–713
data security 300–301, 316, 320, 373, 388, 492
data sharing 295–296, 302–306, 757
data warehouse 297–298, 373, 386
dealer network 583, 671
decision tree 126, 703, 713, 747, 755
deep learning 692, 714
delivery performance 107–108, 531
delivery schedule 451–452
Dell Technologies 51, 61, 104, 242–243, 249, 582
Delphi method 88–89
demand variation 80, 148, 332
demand vector 398 – 400
De Meyer, A. 49
Deming, W. E. 12, 113, 598, 601, 623, 687, 689, 701, 722–723
dependent demand 75, 109, 368
design for manufacturing 114, 575
Design of Experiments 272, 688–694, 707–711,714

Design Structure Matrix 658, 661–663
detail layout 257–258
Dick, P. K. 211
Diebold, J. 197
die exchange *see* changeover; Single-Minute Exchange of Dies
digitalization 36, 377 *see also* digital transformation; digitization
digital transformation 36, 304, 380, 666, 676–677, 682 *see also* digitalization; digitization
digital twin 235, 298
digital yatai 131
digitization 36, 300 *see also* digital transformation; digitalization
DIKW 36
directed putaway 291, 544
disassembly 132–133, 187–191, 193
Discounted Cash Flow 613
discrete manufacturing 76–77, 104, 123–125, 155, 158, 193
dispatch lists 354, 363, 369
divergent processes 121–122
division of labor 25–26, 161
DMAIC 711, 724, 735
DOE *see* Design of Experiments
dock assembly 164
Drucker, P. 29, 46, 591, 598, 600, 603, 622–623
DSM *see* Design Structure Matrix
dual-card system 358
dunnage 351, 447–448, 452, 478–481, 484
DuPont model 597
dynamic allocation 518–520, 530–533, 543
dynamic pull 437, 570
dynamic replenishment 569

E

EEBQ *see* Economic Batch Quantity
e-commerce 87, 299, 384, 493, 582, 584
Economic Batch Quantity 340
Economic Order Quantity 340
economies of scale 25, 58, 66
EDI *see* Electronic Data Interchange
effective capacity 97
effectiveness 3, 8
efficiency 3, 8, 71, 97, 261
EFQM 666
eKanban 563
Electronic Data Interchange 305–306, 335, 370, 494, 570–571
Electronics Manufacturing Services 180
electronic spreadsheet 35, 120, 291, 319–320, 373, 385, 694
electrostatic discharge 261– 262, 448–449, 477
email 295, 301, 319, 322
embedded controller 222, 292, 692
emergencies 82, 379, 575, 624–635, 696
employee store 549
EMS *see* Electronics Manufacturing Services
engineering change 116, 134, 368, 379, 393, 415, 417, 700
Engineering Change Request 394
engineering drawings 27, 34, 225, 688
engineer-to-order 65, 99
Enterprise Resource Planning 35, 290–291, 297, 314–318, 333, 350, 368–373, 378, 395–397, 533
EOQ *see* Economic Order Quantity
equipment master data 410–411
ERP *see* Enterprise Resource Planning
ESD *see* electrostatic discharge
EthonAI 714–715
every-product-every strategy 343–345
Excel *see* electronic spreadsheet
explosion proof 262–263
exponential smoothing 82–84

F

Faber-Castell 39
fabrication processes 113, 126, 133–134
Facilities department 252, 254
Failure Mode and Effects Analysis 756
Faraday cage 448, 477
fast-moving consumer goods 76, 81
Fayol, H. 4
Fazar, W. 659
Feigenbaum, A. 665, 689
Ferdows, K. 49
Ferrari 634
Ferrozzi, C. 568
FIFO *see* first-in-first-out
Fine, C. 54
finished goods inventory 65, 333, 516, 583
Fink, O. 756
fire prevention 258–259, 261, 264, 627–631
firewall 261, 268
first-in-first-out 143, 220, 362–363, 461, 536, 691, 694–695, 716–717
first-pass yield 138, 602, 604, 622
fishbone diagram 655, 728–729
Fisher, R. 707, 709
fixed-position layout 239, 242–244
fixed-station assembly 104, 126, 243
Flexible Manufacturing System 206–208, 268, 350
flow diagrams 274, 277–278
flow rack 182, 268, 345, 351, 362–363, 425, 433–434, 436, 449–450, 508–510, 520, 531, 534–536
flow-shop layout 239–243, 282
flying carpet 185
FMEA *see* Failure Mode and Effects Analysis
FMS *see* Flexible Manufacturing System
FOB *see* free on board
focused factory 67, 280
focus groups 88–89
fool-proofing *see* mistake-proofing
Ford, H. 28–29, 161
Ford Motor Company 10, 28–29, 156, 161, 163, 197–198, 488, 570–571, 634, 724
Ford River Rouge 560, 601, 645
forecasting 71–89, 99–103, 332–335, 569
forecast error 84–85
forklift 425–432, 455, 461–462, 540, 760–762
Forrester, J. W. 12, 118, 565
fourth-party logistics 494
four Ms 9
Foxconn 58–59, 180
FTL *see* full truckload
Francis, C. A. 27
free on board 475
Fujimoto, T. 31, 578, 627–628, 650, 668
Fujitsu-Fanuc 35
Fukushima earthquake 628–630, 632
fulfillment process 498
full truckload 479, 499, 514
functional tests 731, 757
Fung, V. 503

G

Gadani ship-breaking 188
Galton, F. 85
game theory 576, 713
Gantt chart 151, 272, 658
gauge checks *see* go/no-go checking
GE *see* General Electric
gemba 32
genchi genbutsu 656, 672, 727
General Electric 148, 711
General Motors 29, 35, 74, 570–571, 583,
Gilbreth, F. 31, 117, 264
Gilbreth, L. 31, 117
Giorgio Armani 634
GKN Sinter Metals 365–366

global logistics 472–473, 496, 507
Global Trade Item Number 493
GM *see* General Motors
Goldman, S. 60
Goldratt, E. 274, 622, 658, 660
go/no-go checking 115, 210, 690, 695, 706, 716–719
Google 248–249, 493, 598, 707, 752
government inspectors 245, 251
gozinto 391–392, 396–399, 401–402
Gozinto, Z. *see* Vázsonyi, A.
Grande Encyclopédie 22
Grant, R. 57
green corners 251, 265
greenfield 46, 255
Green Performance Map 611
Gribeauval, J.-B. V. de 27
Griffiths, S. 24
Group Technology 280–281
Grove, A. 598, 752
Gryna, F. 601
GTIN *see* Global Trade Item Number
guild 23, 27

H

Hadoop 300
Hall, R. W. 667
handicraft 219, 581
hand tools 25, 178–179, 183–184, 219
Harder, D. 197
Hargreaves, J. 24
Haribo 581
Harry, M. 711
Hayes-Wheelwright product-process matrix 125–126
hazardous material 259, 261, 477, 741, 749
heijunka *see* leveled production
Heinrich's safety pyramid 610
Heinrich, W.H. 610
Herbert Lindner factory 266
Hewlett-Packard 303, 572, 599
HR *see* human resources
hidden factory 195, 227, 232
Hilti 178–179, 244
Hines, P. 671
Hitachi ABB 713–714
HMI *see* Human-Machine Interface
hockey stick effect 622, 624
Holt, S. 54
Honda Motor Company 115, 166, 176, 602, 690, 729
Honda, S. 115
Honest Tea 60–61
Hop, K. 574
hoshin kanri *see* hoshin planning
hoshin planning 599, 678, 723
HP *see* Hewlett-Packard
Hounshell, D. 197
housekeeping 738, 741, 750
Huawei 629
human error 459–460, 575, 691–695, 720
Human-Machine Interface 202, 221–222, 224, 291, 308, 312–314 *see also* operator interface
Human Resources 14, 200, 291, 316, 371, 378, 672–673, 741
Hyatt Roller Bearing Company 127

I

IBM 303, 411
iceberg model 671
ICQCC 725
IE *see* Industrial Engineering
IIoT *see* Industrial internet of things
IKEA 43–45, 75, 133, 513, 517, 524–528, 531, 534, 545–546, 584–585
Imai, M. 641, 645, 724
Immelt, J. 711
improvement project *see* Kaizen Blitz; Kaizen event
inbound logistics 336, 428, 471, 498, 567
incomplete products 337–338, 731
incoterms 475
incremental automation 32, 200, 210
independent demand 75, 109, 368
industrial internet of things 290–291, 294, 308, 688, 712
industrial park 509, 560
industrial revolution 21–27, 156, 161, 196, 666–676
industrial robot 35, 203, 206, 211–212
Industry 4.0 21, 304, 323–325, 546, 666, 676–677
information theory 703
Initial Quality Surveys 690
Injex 516
in-memory database 29
In-plant milk run 427–430
in-process quality 602, 716
input-process-output model 7–8
inspection 132, 139, 191, 688–691, 698–705, 714–733
Intel 56, 116, 249, 416, 598, 751–752
interchangeable parts 27–28, 685, 688–689
intermediate store 549, 552
intermodal containers 626, 473
intermodal transportation 471, 473
Internal Rate of Return 613
International Motor Vehicle Program 667
Internet of things *see* industrial internet of things
inventory accuracy 370, 432, 537, 540, 545–546, 555–556, 608
inventory holding cost 32, 608–609
IoT *see* industrial internet of things
IPO *see* input-process-output
Irkut 248
Iron Fireman 61
IRR *see* Internal Rate of Return
ISA-95 292
Ishikawa's 7 tools of quality control 650
Ishikawa, K. 649–650, 654–665, 689, 694, 724–725
Ishikawa diagram *see* fishbone diagram
ISO 9001 380
ISO standard 277, 303, 473
IT department 14, 291, 316, 321, 323–325
item IDs 401, 403
item master 397, 402
IT/OT stack 34–36, 290–292
IT system 300–301, 314–319, 374

J

Jabil Circuit 58
Jacquard loom 196, 310
Japanese Union of Scientists and Engineers 579, 642, 690, 722
JavaScript Object Notation 293, 306
JD Power 602
Jefferson, T. 27
jidoka 32, 52, 198–199, 668–669, 682
Jingdezhen 22
jishuken 575–576, 579, 642
JIT *see* just-in-time
JKK 697, 726
JMA 667, 753
job instruction 9, 647
job satisfaction 161, 170–171
job-shops 126–129, 169, 239–241, 250, 281–282, 361–362
Jobs, S. 582
Johnson & Johnson 733–734
joint venture 35, 204, 495, 626, 648, 667
Jones, D. T. 33, 668, 673
Jotun 495
JSON *see* JavaScript Object Notation
Junkers aircraft company 106, 669
Juran, J. 685–686, 689,
JUSE *see* Japanese Union of Scientists and Engineers
just-in-time 32, 52, 434, 568, 667–670, 673
just-in-sequence *see* body-on-sequence

K

kaiaku 639
kaikaku 639
kaizen Blitz 642 *see also* kaizen event

kaizen event 33, 576, 642, 652, 655 *see also* Kaizen Blitz
kaizen shop 644
Kalmar factory 164
kanban 334–335, 345, 350–363, 445, 488–489, 506–509, 563, 571, 663–664, 670
kanban board 345, 356–359, 509, 664
Kano model 690
Kano, N. 665, 690
Kaplan, R. 596, 613–614
karakuri 196, 651
Kelley, J. 659
key approach 377, 382–388, 418
Key Performance Indicator 591–593
Kingman's formula 94
kitting 242, 433, 447–448, 453–454, 482, 494, 498, 510
Klann, W. 28
Knuth, D. 36
Kojo Kanri 646
KPI *see* Key Performance Indicator
Krafcik, J. 33
Kumabe, K. 266

L

last-minute customization 485, 494, 498
last-in-first-out 143, 183, 220, 247, 503, 535–536
late-stage customization 495, 511
layout 235–286
layout planning 254–259
Lazy Susan 183
LCL *see* less than container load
lean production 31–34, 49, 667–676 *see also* Toyota Production System
lean six sigma 33, 690, 711
Lean Suppliers Association 575
Lee, H. L. 74, 565
LEGO 301, 402, 723
Lepore, J. 205
less than container load 476
leveled production 99, 192, 336–337, 344, 346–348
life expectancy 686
LIFO *see* last-in-first-out
lights-out strategy 35, 195, 200
Lindner Factory 266
Lindner, H. 266
line assembly *see* assembly line
line balancing *see* leveled production
lineside supply 447–448, 452–454, 466
Ling, D. 101
Little's law 93–94, 445–446, 570, 605–608, 760
loading station 207–208, 229, 536
location address 523–524
location labeling 350–351, 523–524
loss function 709–710
Luddites 25, 43

M

machine controller 302, 309–315, 380, 744
machine learning 82, 712, 757 *see also* artificial intelligence
machine status 366, 377
machine tools 133–134, 141, 200, 213, 309–310, 688, 741
machining cell 141, 147, 152, 208–209
machining center *see* Computer Numerical Controls
MAD *see* Mean Absolute Deviation
maintenance 8, 14–15, 736–767
maintenance cost 760–761
maintenance department 739–744
Maintenance Management Systems 411, 747–748
maintenance outsourcing 745–747
Maison du Gruyère 124
Ma, J. 87
make or buy 58–64
make-to-stock 29, 65–67, 513–514
make-to-order 65–67
Makridakis Competition 82
Makridakis, S. 82
Malcolm Baldridge National Quality Award 665–666, 690
Management by Objectives 598–599, 623
MANUFACTURES 3, 10–11
Manufacturing Execution System 291, 314–316
manufacturing jobs 3, 39–40
manufacturing repurposing 633–634
Manufacturing Resource Planning 35, 303, 316, 368–373
manufacturing strategy 43–69
market intelligence 74, 498, 582
The Markov property 759
Marx, K. 25, 219
mass production 21, 28–29, 668–669
master data 377–419
master data management *see* master data
master data revision management 416–417
Master Production Schedule 72, 103, 329, 335, 349, 374
Materials and Information Flow Analysis 118–119, 563 *see also* Value Stream Mapping
Materials Requirements Planning 35, 303, 329, 368–374, 438
matrix organizations 16
MBO *see* management by objectives
McElroy, J. 601
McLean, M. 472–473
McNamara, R. 29
MDM *see* master data management
Mean Absolute Deviation 85
Mean Forecast Error 84–85
Mean Time Between Failures 757–760, 762
Mean Time To Repair 757–760, 762
Mercier, J.-P. 746
Mercedes Benz 163, 507
MES *see* Manufacturing Execution System
Message Queuing Telemetry Transport 294, 304
Measurement System Analysis 139
Mercedes Benz 507
metal printing 149–150, 263 *see also* additive manufacturing
Methods Time Measurement 176 *see also* MOST
metrics *see* Key Performance Indicators
metric gaming 591, 622–623, 635
MFE *see* Mean Forecast Error
Microsoft Dynamics 317
microstoppages 310, 737, 739–741
MIFA *see* Materials and Information Flow Analysis
milk run 428–431, 445–446, 452, 499–504, 506, 565
Minard, C. J. 119
Minoura, T. 19
mission statement 43–44
Mississippi flood 625–626
mistake-proofing 310, 484, 688–697, 718–721
Mitsubishi 106
mixed-flow assembly 122, 336, 344–347 *see also* mixed-flow line
mixed-flow line 103, 280, 346–347, 444, 454, 733 *see also* mixed-flow assembly
Monden, Y. 348, 667, 718
monument 143, 260, 267–273, 285
morale 116, 170–171, 549, 600, 643, 725
Motorola 666, 694, 311
MPS *see* Master Production Schedule
MQTT *see* Message Queuing Telemetry Transport
MRP *see* Materials Requirement Planning
MRP-II *see* Manufacturing Resource Planning
MTBF *see* Mean Time Between Failures
MTM *see* Methods Time Measurement
MTO *see* make-to-order
MTS *see* make-to-stock
MTTR *see* Mean Time To Repair
multiple sourcing 628
Musk, E. 204, 232
Muther, R. 278

N

nagara switch *see* as-you-go switch
Nagel, R. 666
naïve forecasting 82–83, 444, 569
Nakajima, S. 603, 753
Nalebuff, B. 53, 60, 334, 578
name-value pair 297, 383–385
naming convention 385–389
NASA 12, 658
natural disaster 88, 656
NCR 697
near misses 609
Nespresso 62
neural networks 82, 134
new product introduction 15, 319, 578, 694
Nippon Steel 5
Nissan 33–35, 648
Nisshin Kogyo 631
nomenclature 377–378, 381–390, 416
non-value added 446–447, 477, 545, 606, 622, 699
Norman, D. 225, 719
Norsk Hydro 300–301
Norsk Titanium 149–150
nuclear plant 187–188, 226, 624, 744–746, 755
NUMMI 204, 516, 571, 625–626, 648, 667–668

O

obeya 639, 655–657
Objectives and Key Results 598
obsolescence 29, 452, 496, 609, 694, 730
OEE *see* Overall Equipment Efficiency
OEM *see* Original Equipment Manufacturer
OFAT *see* one-factor-at-a-time
office area 214, 657
off-line rework 732
Ohno, T. 52, 213, 434, 647, 667, 669–670
OKR *see* Objectives and Key Results
Olhager, J. 65
omnichannels 585
Omron 131
Omron Volume-Variety chart 131
one-factor-at-a-time 708
one-piece flow 94–95, 138, 142–143, 208–210, 271, 460, 691, 695, 702, 715–717
one-point lessons 620, 648
OPC-UA standard 304
Operational Excellence 33, 681
operator interface 225 *see also* Human-Machine Interface
operator job design 144, 150–152, 236, 250 *see also* assembly instructions; operator instructions; work combination chart
operator instructions 309, 389, 413, 621 *see also* assembly instructions; operator job design; work combination chart
order-driven activities 65
Order Fulfillment Lead Time 604
organization chart 13–14
OT stack 290–292
outbound logistics 470–471, 498, 562, 567
Overall Equipment Effectiveness 603–604
overhaul 26, 738, 744–745, 755
overproduction 8, 478, 595
overtime 18, 72, 101, 337, 654, 7255

P

Page, L. 752
Paired-cell Overlapping Loops of Cards with Authorization 362–363
palletizing 424, 426, 458–460
pallet exchanger 142, 220–221
pallet jack 426, 439–441, 461–462, 509, 518
pallet pal 450–451
pallet rack 268, 414, 439–441, 518, 524–526, 531–535, 538, 543–544
paper traveler 413, 417, 517
Pareto chart 77, 650 *see also* Parto diagram
Pareto diagram 275 *see also* Pareto chart
part family 281
part number 381–389
part orientation 460–461
part presentation 446–454
part proliferation 630
part-specific dunnage 447, 479, 484
pay-per-build 507, 550, 567
PDCA *see* plan-do-check-act
PDM *see* Product Data Management
pent-up demand 80, 569–570
performance board 615–616, 621
performance charts 619 *see also* performance board
performance measure *see* key performance indicators
periodic maintenance 738, 756
PERT 659–662, 682
PESTEL 43, 50, 57, 68
PFA *see* Production Flow Analysis
phantom subassemblies 403, 406
phishing 300–301
photolithography 56, 108
Picasso, P. 48, 752
pick-and-place robots 147, 212, 214, 484
pick-by-voice 541–542
picking automation 541
picking error 381, 383, 433, 540–541
pick-to-light 413, 541–542
plan-do-check-act 647, 722–726
plan-do-study-act *see* plan-do-check-act
planned maintenance 97, 740–745
planned obsolescence 29, 730
planning horizon 330–331
planning logic 345, 565
Plant Simulation 120, 284
PLC *see* Programmable Logic Controller
PLM *see* Product Lifecycle Management
poka-yoke *see* mistake-proofing
POLCA *see* Paired-cell Overlapping Loop of Cards
Porsche 246, 585
Porter, M. 53, 57
Porter's five forces 43, 50, 53, 68
P-Q analysis 77, 162, 235, 275–276, 285, 435
predictive maintenance 302, 737–739, 756–757, 762
preventive maintenance 739, 744, 746, 753, 756–758, 762
probability theory 686
problem-solving 574–575, 621, 632, 649–655, 722–730
process capability 28, 122, 134–140, 272, 688–695, 715 *see also* capability
process design 113–153, 177, 187, 256, 412
process mapping 117–120
process types 121, 125, 132–133
Product Data Management 371, 386
product development 15, 113–115, 578, 756
product identifier 308, 388, 381–382, 418
production flow analysis 274, 281–282
production planning and control 44, 329–375
Product Lifecycle Management 291, 371, 378
product-quantity analysis *see* P-Q analysis
Programmable Logic Controller 145, 213, 222, 291, 310–315, 367
Program on Vehicle and Mobility Innovation 557
project management 658–664
Ptak, C. 373
public relations 586, 697, 733
publish-and-subscribe model 293–294, 583
pull signal 32, 350–354, 436–446, 499, 552, 569, 571

pull system 346–347, 350–351, 359, 436–445, 561, 569–570
Punitzer, M. 266
purchase order 335, 360, 604
pushcart 161, 207, 430–431, 462–463
push system 350, 432, 436–439

Q

QA *see* quality assurance
QC *see* quality control
QC circle 725
QR code 222, 385, 490
QRM *see* Quick Response Manufacturing
QRQC *see* Quick Response Qualiy Control
quality assurance 210, 291, 510, 547, 578, 701
Quality Circles 85, 579, 642, 654, 725
quality control 132, 579, 649–654, 665, 688–690, 695, 701, 704–705, 715 *see also* statistical quality control; QC circles; Quick Response Quality Control
Quality department 601–602, 696–698, 704, 733
Quality Surveys 690
queueing theory 93, 759–762
Quick Response Manufacturing 665
Quick Response Quality Control 33, 694–695, 697, 722, 726–727, 735

R

R2R *see* Run-by-Run Control
Racing Spirit 115
radio-frequency identification 307, 385–386, 492–493, 523–524, 533, 542
RAMI 4.0 304
random events 92
rapid prototyping 149
Rauwendaal, C. 133
Ravensburger 391
RCM *see* Reliability-Centered Maintenance
Recalls 733–734
recirculating devices 227–228, 232
refurbishing 189 *see also* remanufacturing
regulatory compliance 264, 559
relational database 297, 385
Reliability-Centered Maintenance 737, 745, 754–756
remanufacturing 187–191
Renesas 632
reorder point 34, 305, 363, 368, 437, 441–442
residual variation 71, 78
resourced-based view 57
Respect for Humanity 32, 672
revision management 291, 299, 313, 416–418
rework 694-695, 730–733
RFID *see* radio-frequency identification
risk management 559, 625, 734, 748, 756
River Rouge plant 29, 560, 601, 645
Robotic Process Automation 212
Rockwell Automation 120
Rogers Corporation 381
Rolex 59
Roos, D. 33, 668
root cause analysis 691, 695, 713–714, 722, 728–729
Rosenberg, J. 752
Roser, C. 362
rotary indexing machines 206–207
Rother, M. 563
routing slip 354, 425
RPA *see* Robotic Process Automation
Run-by-Run Control 140
Run-to-Run Control 140
runners 243, 275, 280, 393, 435–436, 612

S

SAAS *see* software as a service
safety stock 500, 514, 543, 551, 625, 628–629
Salah, G. 752
sales campaign 78, 81
Sales & Operations Planning 101–104
sampling *see* acceptance sampling
sandbox 214, 643
sandcone model 49
SAP 298, 305, 316–317, 395, 412–413
Saridis, H. 542
SCADA *see* Supervisory Control and Data Acquisition
scenario planning 88–89, 625, 635
Schmidt, E. 752
Schonberger, R. 667, 683
Scientific Management 30
SCM *see* supply chain management
SCOR framework 559
Scott, R. 211
scrum 663–664
self-inspection 718
semiconductor fabrication 56, 140, 249, 605, 632, 714
service center 747
setup matrix 119, 342–343, 379
Shah, R. 673
Shainin, D. 689
Shannon, C. 36
shelf address 382
Shewhart, W. 688–689, 694, 705
Shih, S. 58, 59
Shingo model 666
Shingo prize 666
Shingo, S. 118, 666, 669–670, 690, 716, 718–719
ship breaking 188
shipping container *see* container
Shook, J. 563, 621, 668
shrink wrapping 547, 550
Siare Engineering International 634
Siemens 251, 284, 312
Siemens Tecnomatix 284
Silver, B. 490
Silver, N. 73
Simchi-Levi, D. 568
Simon, H. 266
simulation model 120, 177, 274, 283–285
simulation software 120, 153, 284
single-deep pallet rack 414, 425, 518, 526, 535–536, 543–544
single-item pallets 452, 547
Single-Minute Exchange of Dies 272–273, 339, 440, 654, 670, 725
single-piece presentation 433, 447–448, 482
six sigma 33, 665–666, 688, 690, 694, 705, 711, 724
Skeen, D. 293
Skinner, W. 43–44, 47–48, 67
slack capacity 337, 739
Sloan, A. P. 29, 74
small-group activities *see* quality circles; QC circles
SMART goals 44, 594–595
Smart Manufacturing 304, 323
smartwatch 224, 366
SMED *see* Single-Minute Exchange of Dies
smiley curve 58–59
Smith, Adam 25–26
Smith, Anthony M. 755
Smith, B. 711
Smith, C. 373
Smith, R. 35
Smyl, S. 82
software-as-a-service 318
S&OP *see* Sales & Operations Planning
spaghetti diagram 253–254, 274, 277, 282, 285
Spanos, C. 140
spare parts management 748–749
SPC *see* Statistical Process Control
Spinning Jenny 24–25
Springfield Model 27–28
sputtering 132, 200, 218, 270
SQCDEP 615, 647
SQL databases 297, 385
Standard Industrial Classification 6

standardization 66, 231, 273, 303, 668
standards 646–648
Standard Work 152
Statistical Design of Experiments *see* Design of Experiments
Statistical Process Control 28, 140, 299, 688–689, 694, 705–706
Steiff 59
storage devices 268, 517, 555, 609
strangers 129, 134, 244, 275, 435–436, 612
strategy deployment 599, 621 *see also* hoshin planning
subassembly feeder lines 165
subassembly structure 159, 176, 393, 396, 402, 408
successive inspection 690, 716, 718–719, 732–733
suggestion system 652–653
supermarket 414, 424–425, 430–435, 541–542, 549–552
Supervisory Control and Data Acquisition 221–222, 291, 295–297, 308, 312–315, 694, 729
supplier kanban 353–354, 360
supplier milk run 479–480, 499–504
supply chain coordination 561, 634
supply chain management 74, 389, 469, 557–588, 630
supply networks 501, 557, 561, 564, 633
Suri, R. 363, 665
SWOT analysis 43, 50, 55, 57, 68, 649
Synchronous Manufacturing 33
systems thinking 11–13

T

Taber, G. 54
Taguchi, G. 689, 709–710
Taguchi's quality loss function 710
Taiho Kogyo 631
takt-driven production 105, 107–108, 243, 670
takt time 71, 104–109, 144–145, 161–162, 192, 272, 461, 732–733
taxonomy of manufacturing 6
Taylor, F. 30–31
Technical Data Management 8, 417, 612
technology watch 321, 676
Tesla 129, 163, 169, 203–204, 634
TEU 474–475
Texas Instruments 215
Theory of Constraints 274
third-party logistics 494–498, 507
Thoratec Corporation 282–283
throughput time 13, 28, 92, 441
Thurer, M. 362
Thyssen Krupp 495–496
Timbuk2 Designs 104, 175
TOC *see* Theory of Constraints
Toffler, A. 677
tolerance interval 134–137, 710
tombstone 207–208, 229
tool logistics 231
Total Preventive Maintenance *see* Total Productive Maintenance
Total Productive Maintenance 603, 665, 673, 737, 753, 762
Total Productivity Maintenance *see* Total Productive Maintenance
Total Quality Control 665, 688–689
Total Quality Management 665, 673, 688, 690
tow cart 430, 463, 534
tower light 224, 364–365, 374, 410 *see also* andon light
Toyoda, A. 74, 733
Toyoda, E. 19, 202, 670
Toyoda, K. 52, 353, 669–670
Toyoda, S. 27, 52, 74, 199, 669
Toyota 17, 19, 31–35, 49, 52, 74, 106, 199, 213, 256, 272–273, 344, 353–354, 357, 516, 563, 570–571, 578–579, 582, 599, 601, 618, 621, 626, 629– 632, 651, 667–672
Toyota Production System 21, 31–35, 49, 52, 106, 579, 665, 667–671, 674
Toyota's 7 steps of automation for machining 213–214
The Toyota Way 651, 672
Tozawa, B. 641
TPM *see* Total Productive Maintenance
T-point 66
TPS *see* Toyota Production System
TQC *see* total quality control
TQM *see* total quality management
traceability 35, 296, 356, 413, 490, 545, 712
trade-off model 48, 52
Training Within Industry 648, 689
Transaction Cost Theory 63, 68
transfer line 126, 197, 206–207
traveler 126, 240, 413, 488–489
traveling salesrep problem 341
trend projection 82–83
TSMC 113
T-type specialists 20
Tufte, E. 150–151
tugger train 410, 426, 463, 508–510
Tulip 318
Turing, A. 324
TWI *see* Training Within Industry
two-bin systems 329, 351–352, 374

U

U-cell 173, 241 *see also* U-shaped cell
Union Pacific 626
unions 198
United Airlines 754, 756
UPC 493
USA principle 213
usability engineering 195, 224–225, 541, 688, 719–720
US–China trade war 624, 629
U-shaped assembly 170, 183 *see also* cell assembly
utilization 92, 94, 977

V

Value Stream Mapping 34, 118–119, 563
variability 76–80, 96–97, 190–191, 332, 709
variant explosion point 100
VATI analysis 274
Vázsonyi, A. 396
Vázsonyi procedure *see* gozinto
VDA *see* Verband der Automobilindustrie
VDI *see* Verein Deutscher Ingenieure
vendor-managed inventory 74, 83, 370, 435, 567–568, 587
ventilation 129, 261, 267
Verband der Automobilindustrie 305
Verein Deutscher Ingenieure 315
Verity 546
Verschlimmbesserung 639
vertical disintegration 560
vertical integration 29, 560
Vincenti-Brown, C. 666
virtual dualization 627–628
virtual reality 177, 283, 648, 677
visitor center 245–246
visual controls 142, 269, 273, 694 *see also* visual management
visual management 364–367 *see also* visual controls
VMI *see* vendor-managed inventory
Volkswagen 28–29, 166, 389, 601, 669
Volvo 164, 181, 357, 644, 651, 674
Volvo experiments 164
Vonnegut, K. 211
VR *see* Virtual Reality
VSM *see* Value Stream Mapping
VW *see* Volkswagen

W

Wald, A. 689, 694
Walker, M. 659
Warehouse Management System 35, 291, 314, 370, 414, 425, 520, 529–533, 544–546
warranty 730

Warren, M. 156
war room 631, 655, 657
waste disposal 255, 267–271, 612
water spider 425, 430–431, 433–434, 465, 550
Watt, J. 24, 196
WCM *see* World Class Manufacturing
Weiszfeld, E. *see* Vázsonyi, A.
Welch, J. 711
Western Electric 689, 694
wholesale 447, 493, 531, 584
Wi-Fi 293
Wilhelmsen 495–496
Wisdom 36, 651, 672
Wistron 180
WMS *see* Warehouse Management System
Womack, J. 33, 668, 673
work clothes 251, 265, 266
work combination chart 146, 151–152, 175
work combo *see* work combination chart
work instructions 35, 152, 177, 179, 380, 390, 621
World Class Manufacturing 19, 33, 666

X

XML messages 293, 305–306, 334
X-mR chart 705–706
XPS *see* company-specific production system
XYZ classification 275

Y

Yamashina, H. 666
yamazumi chart 192
YAML 293, 306
Yawata Steel Works 5
Ybry chart 151, 272, 362
yearly maintenance 103, 760–762
yield 116, 134, 137–139, 602–604, 692, 694, 707, 713–714
Yoshino, I. 599

Z

Zenbara 176, *177*, 406
Zinkgraf, S. 711
zone identification 527–528
zoning 479, 532